DOCUMENT RESUME

ED 178 685 CE 019 280

AUTHOR Strang, Ernest W.; And Others
TITLE Technical Assistance in Evaluating Career Education Projects. Final Report. Volume IV: Turnkey Manual for Evaluating Local Career Education Projects.
INSTITUTION NTS Research Corp., Durham, N.C.
SPONS AGENCY Office of Career Education (DHEW/OE), Washington, D.C.
PUB DATE Jun 78
CONTRACT 300-76-0312
NOTE 128p.; For related documents see CE 019 277-281

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS *Administration; Administrator Guides; *Career Education; Elementary Secondary Education; *Evaluation; *Evaluation Methods; Field Studies; *Guidelines; *Models; Planning; Technical Assistance

ABSTRACT

This document contains the fourth of five volumes reporting the activities and results of a career education evaluation project conducted to accomplish the following two objectives: (1) to improve the quality of evaluations by career education projects funded by the United States Office of Career Education (OCE) through the provision of technical assistance to fifteen of these projects; and (2) to identify, as a result of this technical assistance effort, a field tested model which can be used by other career education projects funded by OCE in future evaluation efforts. Intended for distribution to participating projects, this volume contains the necessary information for longitudinal continuation of this evaluation at the project level. It is divided into the following sixteen chapters which present a step-by-step procedure for building a comprehensive evaluation design: (1) review program design, objectives, and rationale; (2) specify information requirements and management needs; (3) identify conceptual framework for evaluation; (4) specify evaluation questions of interest; (5) specify data base structure and contents; (6) select instrumentation; (7-10) specify population selection procedures, data analysis techniques, data collection and reduction procedures, and reporting requirements and procedures; (11) identify major assumptions and restraints; (12) define support requirements; (13) develop schedule of evaluation activities; (14) list budget requirements; (15) develop management plan; and (16) prepare the evaluation design document. (BM)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED178685

JUNE 1978

# TECHNICAL ASSISTANCE IN EVALUATING CAREER EDUCATION PROJECTS

Contract 300760312

## FINAL REPORT
## VOLUME IV
## TURNKEY MANUAL FOR EVALUATING LOCAL CAREER EDUCATION PROJECTS

Ernest W. Strang
Helmuts Feifs
Richard J. Rohlf

Submitted to.

DHEW/US Office of Education
Office of Career Education

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

NTS RESEARCH CORPORATION

The material in this publication was prepared pursuant to a grant or contract from the Office of Education, U. S. Department of Health, Education, and Welfare. However, points of view or opinions expressed do not necessarily represent policies or positions of the Office of Education.

DISCRIMINATION PROHIBITED

Title VI of the Civil Rights Act of 1964 states: "No person in the United States shall, on the ground of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance." Title IX of the Education Amendments of 1972, Public Law 92-318, states: "No person in the United States shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity receiving Federal financial assistance." Therefore, career education projects supported under Sections 402 and 406 of the Education Amendments of 1974, like every program or activity receiving financial assistance from the U. S. Department of Health, Education, and Welfare, must be operated in compliance with these laws.

## PREFACE

This report has been prepared pursuant to the requirements of DHEW/USOE Contract Number 300760312, *Technical Assistance in Evaluating Career Education Projects*, for the Office of Career Education. This contract was awarded to National Testing Service, Inc., (NTS) on June 21, 1976, to accomplish the following two objectives:

1. To improve the quality of evaluations conducted by OCE-funded career education projects through the provision of technical assistance to fifteen of these projects.

2. To identify, as a result of this technical assistance effort, a field tested model which could be used by other career education projects funded by OCE in future evaluation efforts.

Contractors are encouraged to provide objective, independent evaluations; therefore, none of the results or positions presented in this report necessarily reflect official Office of Career Education policy.

This is Volume IV of the five volumes comprising the common Career Education Evaluation Final Report. The volumes include the following:

Volume I: Summary Volume. This volume contains highlights of the study design and results.

Volume II: The Career Education Evaluation Final Report. This, the main body of the report, contains detailed information about the background of the evaluation, the evaluation design, and the results of the evaluation. Three appendices are included, as well: (1) summaries of the fifteen local evaluations; (2) a summary of the technical assistance provided by NTS; and (3) all instruments used in the study.

Volume III: The Career Education Instrumentation Substudy. This volume contains the rationale, procedures, and results of the outcome and process instrument validation and construct definition activities undertaken by NTS under this contract.

Volume IV: Turnkey Manual for Evaluating Local Career Education Projects. The Turnkey Manual was prepared for distribution to participating projects and contains the necessary information for longitudinal continuation of this evaluation at the project level.

Volume V: The Career Choice Substudy. This volume contains the design and results of a study of the relationships of career education to secondary students' career choices.

In addition to these five common report volumes, NTS prepared unique local evaluation reports for each of the fifteen participating projects.

TABLE OF CONTENTS


ACKNOWLEDGEMENTS . . . . . iii
LIST OF TABLES . . . . . iv
LIST OF FIGURES . . . . . v
INTRODUCTION . . . . . 1
Use of this Turnkey Manual . . . . . 1
The Need for Evaluation in Career Education . . . . . 2
Defining Evaluation . . . . . 6
Information Based Evaluation . . . . . 6
Limitations of Evaluation . . . . . 8
Longitudinal Evaluation Design . . . . . 11
An Evaluation Design Procedure . . . . . 12
CHAPTER 1: Review Program Design, Objectives, and Rationale . . 1.1
CHAPTER 2: Specify Information Requirements and Management Needs . . . . . 2.1
CHAPTER 3: Identify Conceptual Framework for Evaluation . . . . 3.1
CHAPTER 4: Specify Evaluation Questions of Interest . . . . . 4.1
CHAPTER 5: Specify Data Base Structure and Contents . . . . . 5.1
CHAPTER 6: Select Instrumentation . . . . . 6.1
Steps in Selecting Instrumentation . . . . . 6.3
Instruments in the Longitudinal Design . . . . . 6.14
CHAPTER 7: Specify Population Selection Procedures . . . . . 7.1
Types of Longitudinal Designs . . . . . 7.2
Following-up Your Respondents . . . . . 7.4
Expanding the Size of Your Sample . . . . . 7.7
Comparing Input Characteristics . . . . . 7.8
Control Groups. . . . . 7.9
CHAPTER 8: Specify Data Analysis Techniques . . . . . 8.1
Questions of Description . . . . . 8.2
Questions of Comparison . . . . . 8.4
Questions of Relationships . . . . . 8.11
Questions of Prediction . . . . . 8.13
Predicting to a Criterion . . . . . 8.14
Causal Analysis . . . . . 8.15
Trends Over Time . . . . . 8.17
CHAPTER 9: Specify Data Collection and Reduction Procedures . . 9.1
CHAPTER 10: Specify Reporting Requirements and Procedures . . . 10.1
Special Concerns . . . . . 10.1
Specifying Report Requirements . . . . . 10.2
CHAPTER 11: Identify Major Assumptions and Restraints . . . . . 11.1
CHAPTER 12: Define Support Requirements . . . . . 12.1
CHAPTER 13: Develop Schedule of Evaluation Activities . . . . . 13.1
CHAPTER 14: List Budget Requirements . . . . . 14.1
CHAPTER 15: Develop Management Plan . . . . . 15.1
CHAPTER 16: Prepare the Evaluation Design Document . . . . . 16.1
Preliminary Materials . . . . . 16.2
Chapter Contents . . . . . 16.4

APPENDIX: Evaluation Design Checklist . . . . . A.1

## Acknowledgements

Rarely can the authors of documents of this size claim sole ownership; this is not an exception. Indeed the job of "author" has often taken the form of compiling and editing the many excellent suggestions we have received.

We would like to express our foremost appreciation to Mr. Terry Newell, our OCE project officer, for his extensive review efforts and, perhaps more importantly, for his support and gentle prodding throughout.

The directors and their staffs of these local career education projects also earned our sincere thanks. Nancy Losekamp, Tony Cavallo, and Lloyd Swanson reviewed our early efforts and provided the constructive criticism necessary to get us "on track."

A large number of NTS staff contributed to this Manual. A. Jackson Stenner graciously allowed us to use the IBE Manual as a starting point and suffered through a number of early drafts. Joan Troy also read our drafts and was able to offer a great deal of assistance. Robert Pinnell was able to make our inept prose readable. Pauline Matson, Ray Alderson, Phyllis Thomas, and Annie Mae Price fought through our handwriting to type successive drafts. We extend our sincere appreciation to all those involved in preparing this Turnkey Manual.

Ernest W. Strang
Helmuts A. Feifs
Richard J. Rohlf

## LIST OF TABLES


| Table | | Page |
|---|---|---|
| 1 | Table of Correlations Among Self Concept and Achievement Scores in Basic Skills for a Sample of Fifth Graders. . . . . | 8.16 |

## LIST OF FIGURES


| Figure | | Page |
|---|---|---|
| 1 | Flow Chart for Developing an IBE Evaluation. . . | 14 |
| 2 | Example of a Completed Domain Analysis Matrix for a Career Education Program . . . . . | 3.3 |
| 3 | Example of a "Program Requirements Matrix" Developed from the Domain Analysis Matrix in Figure 2. . . . . | 3.4 |
| 4 | Adjusted Posttest Outcome Profiles for Grade 5 Students Receiving High or Low Exposure to Career Education . . . . . | 8.3 |
| 5 | Correlation Between Basic Skills Achievement and Selected Career Education Dimensions . . . . | 8.5 |
| 6 | Exposure Index Scores Regressed on Adjusted Career Awareness Posttest Scores . . . . . | 8.10 |
| 7 | Balantine for Process Dimensions and Group Membership Predicting Residual Outcomes. . . . . | 8.12 |
| 8 | Graphic Presentation of Cross-lagged Panel Data from Table 1. . . . . | 8.16 |
| 9 | Career Awareness Scores (Constant Improvement) . | 8.18 |
| 10 | Career Awareness Scores (Variable Improvement) . | 8.20 |
| 11 | Flow Chart for Collection of Input, Process, and Outcome Data . . . . . | 9.2 |
| 12 | Table of Contents for an Evaluation Design . . . | 16.3 |
| 13 | Evaluation Design Checklist. . . . . | A.2 |

# INTRODUCTION

## Use of This Turnkey Manual

No "best" way to do an evaluation exists, whether the evaluation concerns career education or any other educational program. Nonetheless, there are some basic principles and practices which will enable you to obtain the most useful information from whichever evaluation design you choose to use. It is the hope of those of us at NTS who have been working with your projects that the suggestions offered in this Manual will allow you to continue to carry out your evaluations with little if any outside help.

More specifically, when used with the Evaluation Design Document, the Common Evaluation Report, and the Report on your project prepared by NTS as part of Technical Assistance for Career Education Evaluation (OCE contract #300-76-0312), this Manual will permit continuation and completion of the longitudinal evaluation of your project's career education efforts begun during the 1976-77 school year. It is important to note that this Manual does not, and is not intended to, stand alone. Given the complexities involved in evaluation generally and career education in particular, a single inclusive Manual would be of overwhelming bulk. This Manual, however, will provide you a brief

summary of the evaluation process including some of the potential pitfalls and problem areas, as well as tips and suggestions for avoiding or overcoming them.

Particular attention will be given to assisting you in developing and specifying your evaluation questions of interest, specifying the appropriate data base structure and contents, instrument selection, sampling techniques, and the collection, reduction and analysis of your data. The brief summaries of these and the other topics included in this manual can be expanded by you to meet your own particular needs by referring to the relevant sections in the documents mentioned above.

## The Need For Evaluation in Career Education

Career education, as a distinct term and an Office of Education initiative, began in 1971. Between 1971 and 1974, the concept of career education was developed through a series of projects across the nation, supported primarily with vocational education funds. In 1974, Congress officially endorsed career education through the passage of Section 406 of the Education Amendments of 1974. In that legislation, Congress called for increased dialogue concerning the concept of career education as well as demonstration that career education could become an effective program. The Office of Career Education (OCE) in U.S.O.E. was established by Congress to accomplish these goals. OCE has funded projects in FY75 and FY76 to meet the congressional mandate.

Because of this charge to demonstrate effective career education, OCE has recognized the necessity for devoting considerable attention to

the evaluation of career education. This concern for evaluation has also emerged from the GAO study of career education released in January 1976.[1]

The evaluation of career education is, however, still in its infancy. While a variety of instruments, designs, and techniques have been tried, no clear consensus exists on the best approach to take. There is a need for the development of workable evaluation techniques. There is also a need for providing training and assistance to funded career education projects to enable them to mount successful evaluations. While it may not be traditional in the social sciences to try to solve evaluation problems while the program concept and practices are still being developed, the demands being made for educational accountability require that evaluation be tackled now in career education.

In the first year of its operation, OCE tried to be responsive to evaluation needs by emphasizing the importance of evaluation to the forty-five local school system projects it supported. Project staff, however, were unable to perform sophisticated evaluation both because they lacked the skills and funds to do so and because reliable evaluation techniques were not available to them.

To avoid a similar problem during the second and future years of its operation, OCE awarded a contract to National Testing Service (NTS) in late June, 1976, to accomplish two major objectives:

---

[1] "Career Education: Status and Needed Improvements," U.S. General Accounting Office, January 26, 1976.

I. To improve the quality of evaluations conducted by OCE-funded career education projects through the provision of intensive technical assistance to fifteen of these projects; and

II. To identify, as a result of this technical assistance effort, a field tested model which could be used by other career education projects funded by OCE in the future.

In an address delivered at the Commissioner's National Conference on Career Education on November 8, 1976, Dr. Sidney P. Marland, Jr. placed the need for evaluation firmly within his set of "imperatives" for career education:

> We must assess the outcomes of career education and document the evidence that, where career education has been systematically installed, the academic growth of young people, including those in college, improves dramatically, with intrinsic motivation deriving from the learners own reasoned application to the learning task. Most of us have been too busy implementing career education to take the pains to document for others the measureable gains.[2]

In testimony prepared for presentation before the subcommittee on Elementary, Secondary, and Vocational Education, United States House of Representatives on February 9, 1977, Kenneth B. Hoyt, Director of the Office of Career Education, noted that "Evaluations of effectiveness for career education have been generally positive. Yet much remains to be done before we will be able to claim that

---

[2] "Career Education Update," National Advisory Council for Career Education, Washington, D. C., 1977, p. 15.

career education has been fully validated."[3]

This comment reiterated Hoyt's longstanding commitment to the need for evaluation. Speaking before the National Advisory Council for Career Education in July, 1976, he made the following points:

> Wherever comprehensive career education efforts now exist, there is a critical current need to engage in systematic evaluation of their effectiveness. Two factors have delayed this effort to date. First, many have pointed out that, because it is a developmental concept, the truly important long run effects of career education cannot be known for at least another 20 to 30 years. Second, much attention has been devoted to our lack of suitable instrumentation, in such domains as teacher/student attitudes, work values, decision-making, and self concept, that is currently hindering our efforts aimed at short-term evaluation of career education's effectiveness. I am convinced that, while both points are valid and important, neither can be allowed to serve as an excuse for failing to greatly increase our attempts to evaluate current comprehensive career education efforts wherever they can be found.
>
> As a vehicle for reform of Education, career education should not be afraid to use traditional evaluative criteria - including increases in pupil achievement, in pupil attendance, and in school holding power. Of these, demonstrations of the ability of a career education approach to increase pupil achievement in the basic academic skills is, by far, the most crucial. While preliminary evidence available suggests that career education can have such an effect, this evidence is, at present, far too sparse and too susceptible to criticism for us to claim that career education, if properly applied, will produce such results. With the current strong national push toward a "back to basics" emphasis, it seems to me especially crucial that comprehensive career education efforts demonstrate that, in fact, they can serve as an effective vehicle for improving basic achievement.[4]

---

[3]"Testimony Prepared for Presentation before the subcommittee on Elementary, Secondary, and Vocational Education, United States House of Representatives," February 9, 1977, p. 12. (mimeo)

[4]Kenneth B. Hoyt, "Next Steps for Career Education," Remarks Prepared for Presentation to the National Advisory Council for Career Education, Washington, D. C., July 28, 1976, pp. 4-5. (mimeo)

## Defining Evaluation

There is a clear and urgent need for evaluating the processes and outcomes of career education programs. "Evaluation," just as has been the case for "career education," has been given a number of meanings. Before entering a discussion of evaluation designs, it is first necessary to describe "evaluation" as used in this particular effort. The expanding literature on educational evaluation indicates a transfer of attention from philosophical discourses on the need and desirability of educational evaluation to the critical appraisal of the tools and methodologies available. This shift in discussion has paralleled the realization that evaluation can be an aid to rational thought and action within the decisionmaking process. In order to approach the study of educational evaluation in a systematic manner, we must conduct a more thorough examination of its components and elements than is generally found in the literature.

Evaluation includes specifying an information need and collecting, analyzing, and reporting information to alleviate that need. Most information needs are generated by individuals in decisionmaking capacities. In education this includes, among others, teachers, principals, superintendents, board members, state education agency staff, and legislators. Each of these groups or individuals makes decisions at different levels and, thus, requires information in different formats and at different times.

## Information Based Evaluation

Evaluation is a continuous process involving the constant reappraisal of approaches and policies so that rational decisions

can be made among often conflicting alternatives. Evaluation is not a product, but a means for enhancing the use of available resources; it provides the information required for planning and managing an enterprise. Educational evaluation can be usefully defined as follows:

> Evaluation is the process of clarifying a set of information needs, and collecting, analyzing, and reporting information to alleviate those needs.

This definition includes several distinct elements. A short discussion of these elements will serve to clarify the meaning of the above definition and present some of the topics which will be expanded in later sections.

1. Evaluation is the Process...a continuous activity requiring an expenditure of resources in order to be sustained. Evaluation as a process should be distinguished from a "plan." A "plan" is a set of decisions impacting on the future, which can be developed through evaluation feedback or through some other, perhaps "irrational," method of decisionmaking.

2. Of Clarifying...the evaluation process seldom includes the definition of information needs. This is the province of management. The role of evaluation is to clarify these identified information needs in such a manner that valid data can be secured to alleviate the needs.

3. A Set of Information Needs...information needs seldom are singular in nature. More often, a decisionmaker requires information on a number of variables, many of which are interrelated. "A set of information needs" then, refers to a systematically interrelated matrix of choices among alternatives.

4. And Collecting...the collection of information which will have impact on the decisionmaking process involves an economically selective approach. Rarely is there no information available that will effect a decision making situation. Generally, there is an abundance of information, and the problem becomes one of maximizing the amount and quality of the information collected within fixed resource constraints.

5. Analyzing...analysis involves the manipulation of the collected information in such a way that the decision needs can be best alleviated. Just as the collection element involves a number of options and requires selectivity, so does the analysis element. More commonly, evaluation data is manipulated too little rather than too much.

6. And Reporting Information...All previous elements are meaningless academic exercises unless the results of the evaluation process can be communicated to the right decisionmakers in a timely, precise, and readily understandable format and language.

7. To Alleviate Those Needs...the alleviation of need implies that the preceding elements of the definition have been adequately realized. As previously mentioned, evaluation is not an end in itself; the end, rather, is the provision of accurate information to decisionmakers. The decisionmakers solicit this information in the hope that an adequate knowledge base can be realized.

## Limitations of Evaluation

Evaluation is not without deficiencies. The concept is in its infancy. Some of the limitations outlined below are inherent in the

way an evaluation is implemented; others are more the result of immaturity than of any inherent deficiency.

## Evaluation is Incomplete

The two most obvious reasons for the incomplete nature of evaluation are resource limitations and constraints imposed by the decisionmaking cycle. It is a rare occasion when sufficient money and personnel are available to investigate a problem adequately. Likewise, it is a rare decision or decisionmaker that can wait indefinitely for an information base. The need for reporting almost always limits the scope and comprehensiveness of the evaluation.

Another reality is that, no matter how large the budget is for an evaluation study, the evaluators can never collect and analyze all the information which might be relevant. One very important skill for an evaluator is the ability to identify those information sources which can be investigated (within budget constraints) and will have the greatest impact on decisionmaking.

## Measures are Limited

In education, measures of effectiveness are, at best, reasonable approximations of often vaguely defined outcomes. Generally, the best that can be accomplished is to select measures which at least appear to sample the behavior of interest. Consider "self-concept," for example: it exists only in the mind--and in the student's mind at that. We cannot, therefore, measure the effectiveness of programs aimed at improving the "self-concept," but must use instead such approximation as attendance, observable behaviors, and attitude inventories of often limited applicability.

The Uncertain Future

Planning and decisionmaking in education have always been more an art than a science; what appears to be happening is that the basis of the form is changing from a crisis, ad hoc approach, based purely on intuition and judgment, to one based on intuition and judgment supported by empirically based information. The change is a slow but apparently inevitable process. With this growth comes skepticism, some of which appears justified, since evaluation is not always done competently and evaluators have not always openly expressed inevitable limitations. There may indeed be dangers in using evaluation where broad educational policy decisions are at issue. For one thing, many factors important in education are not readily amenable to quantitative analysis; as such they may be neglected, set aside for further consideration, and then forgotten, or more likely, improperly weighted in either the evaluation or decisionmaking process. In other instances, an evaluation study may, on the surface, appear so scientific and quantitative that it may be assigned undeserved merit because of the many subjective judgments involved. In other words, we may be so mesmerized by the apparent precision and conclusiveness of the numbers that we overlook the simplifications made to achieve this precision and, thus, neglect important qualitative factors. However, after all these limitations are said and done (though they should not be ignored), it is apparent that the alternatives to evaluation (crisis management and seat-of-the-pants decisionmaking) are even more riddled with dangers in the neglect of important considerations and assignment of improper weights to some factors.

## Acceptance of Evaluative Information

Unfortunately, evaluative information is not always an unmixed blessing for either the program staff or the decisionmakers. The presentation of information which indicates that an alternative or program is not effective is usually insufficient. Even when the information is relatively conclusive, it is difficult to convince committed educators that a pet program or activity should be reduced, severely modified, or eliminated.

## Longitudinal Evaluation Designs

The central purpose of this manual is to assist you and administrators of the other fourteen projects to continue the implementation of some or all aspects of the design developed by NTS. The importance of collecting longitudinal data is clear. As was mentioned in a previous statement by Kenneth B. Hoyt (see page 5), it is possible that "the truly important long range effects of career education cannot be known for at least another 20 or 30 years." NTS's design was not, of course, developed to cover a span of that length; nonetheless, an additional one or two year effort in those areas seen as particularly important to you will greatly assist in your future program development. Perhaps the greatest mistake in evaluating career education, of which for a time NTS was guilty, was the assumption that a pre-post design must start and stop at ten-month intervals. What, in retrospect, seems like a simple notion now looms as what NTS perceives to be the largest single contribution we have made to several of the career education programs we currently evaluate--namely, the longitudinal (in some cases, three-year) evaluation design. Such a design poses special problems of student attrition

and noncomparability of test levels, but the payoff promises to be substantial. It appears that fond hopes of program effects being measurable over a one-year period are folly--attitudes and behaviors that are important take more time to influence than can be given in one school year. Nothing about the longitudinal design precludes yearly pre-post testing; in fact, combining a longitudinal design with pre-post, one-year designs is a cost effective strategy if judicious use is made of sampling techniques.

Although a number of suggestions and general guidelines will be provided in these pages, the most important is that for your evaluation results to be useful you must first pose the right evaluation questions of the right sources. All else is secondary.

## An Evaluation Design Procedure

The procedure presented in the following pages relies heavily on the interaction between the evaluation team and potential information users, including program staff. Perhaps the most effective technique for facilitating this interaction is the use of a design conference. At the conference relevant information users can begin to identify their information needs. If certain potential users cannot be present, a role-playing technique can be used quite effectively to project their needs. Whenever possible the information user should be polled in person. This results in a more accurate picture of his needs and provides the user with a sense of ownership in the evaluation process.

The design conference should be scheduled as soon after the initiation of project activities as possible, as the topics it should address include the steps involved in activities 1 through 4.

The flow chart presented in Figure 1 provides a logical step-by-step procedure for building a comprehensive evaluation design. Each step or activity in the flow chart results in a product which becomes the foundation for subsequent evaluation activities. The final product of these activites is a blueprint or plan for conducting the evaluation. This plan is generally presented in the form of a design document or section of a proposal, and the product of each activity in the flow chart becomes a chapter in the final plan.

Each cell in the flow chart is a definable activity with a unique product; none of the activities and few of the products have meaning except in relation to the other products. As an example, suppose the following evaluation question is posed: "Do students with exposure to career education do better on a standardized reading achievement test than a group of like students with no such exposure?" This evaluation question becomes the basis for all subsequent activities in the flow chart. A relevant data base is established; an appropriate instrument is selected; the proper analysis technique is selected; specific data collection procedures are established; and so on through the rest of the flow chart activities. To illustrate the interdependence of activities, consider the fact that the analysis technique can not be selected until the evaluation question has been formulated, and the data base elements have been established. The flow of activities is quite straightforward and understandable; however, an understanding of the interdependence and inter-relatedness of the activities is crucial to using the flow chart to design evaluations

## FLOW CHART FOR DEVELOPING AN IBE EVALUATION

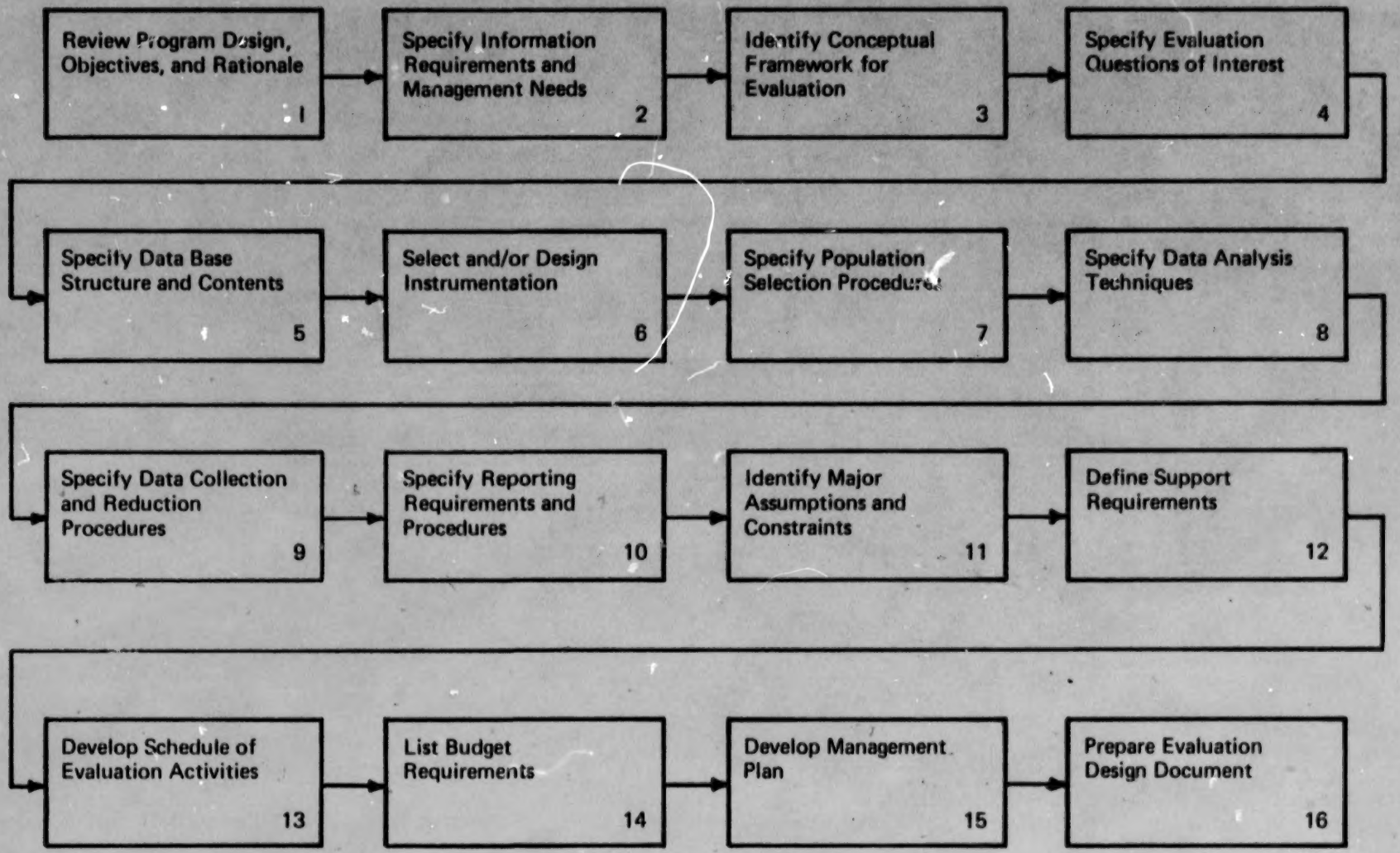

Review Program Design, Objectives, and Rationale 1
Specify Information Requirements and Management Needs 2
Identify Conceptual Framework for Evaluation 3
Specify Evaluation Questions of Interest 4
Specify Data Base Structure and Contents 5
Select and/or Design Instrumentation 6
Specify Population Selection Procedures 7
Specify Data Analysis Techniques 8
Specify Data Collection and Reduction Procedures 9
Specify Reporting Requirements and Procedures 10
Identify Major Assumptions and Constraints 11
Define Support Requirements 12
Develop Schedule of Evaluation Activities 13
List Budget Requirements 14
Develop Management Plan 15
Prepare Evaluation Design Document 16

FIGURE 1

In addition, the procedure is iterative in that decisions are made in later activities which require the modification of decisions made in earlier activities.

An evaluation design checklist is included at the end of this manual to help you monitor your progress through this process.

# 1

## REVIEW PROGRAM DESIGN, OBJECTIVES, AND RATIONALE

The intent and focus of this first activity is to conduct a thorough review of the program. It is frequently the case, that this initial activity is thought of in retrospect as being one of the most valuable steps of the entire evaluation sequence. Because you, as project staff, have a vested interest in the program, it is strongly recommended that this program review be accomplished with the assistance of one or more individuals who can bring to this task an attitude of neutrality if not a degree of skepticism. This second party can then assume the role of critic, thus fostering a re-examination of all assumptions and rationale with the degree of clarity needed to assure that the program is internally consistent and externally relevant.

Of particular importance is a rigorous review of the identified needs, the program objectives for meeting those needs, and the program strategies for achieving the stated objectives. This process often results in the tightening of objectives, a restating of rationale, and/or a refocusing of program procedures. The design conference suggested in the Introduction to this Manual is an appropriate arena for completing this activity.

Several suggested process steps for completing this activity include the following:

a. Review rationale (or needs assessment).

b. Establish congruence between program objectives and program rationale.

c. Establish that program objectives are operationally stated.

d. Establish congruence between program objectives and program activities.

The product of these process steps is a consistent program designed to meet identified needs, well developed program objectives, and clearly formulated program procedures. In addition, a concise program description should be prepared.

# 2

## SPECIFY INFORMATION REQUIREMENTS AND MANAGEMENT NEEDS

The range of potential information users and the variety of information potentially available, relative to any given educational program, is virtually boundless. Undoubtedly, this is true for your program. Recognizing that it would be impractical if not impossible to try to provide every possible user with all the conceivable information that they might request, a technique is required for defining the parameters of the information to be gathered or generated.

The two most logical focal points for such a procedure are the people who desire some information about your program and the types of information they desire. Capitalizing on these two reference points, a technique called domain analysis can be used to define and focus the direction of your evaluation. Domain analysis can best be described by presenting the procedures which comprise the technique:

1. A list of all potential information users is developed. This list might include teachers, district superintendent, school board, state department of education, and USOCE.

2. A second list is made of all potential domains (classes) of information. The list might include such domains of student career education outcomes (test scores), program related costs, staff attitudes, and community acceptance.

3. A matrix is created contrasting the information domains against the information users. (See Figure 2).

4. The matrix is then analyzed to identify the essential information users and information domains.

5. A new, somewhat constricted matrix is constructed summarizing the most important relationships and becomes the basis for subsequent activities. The final matrix is termed the "program information requirements matrix." (See Figure 3).

These activites provide the opportunity for considerable imagination and creativity on the part of the project staff. The comprehensiveness and relevance of the "program information requirements matrix" dictate the parameters for all your subsequent evaluation activities.

The process steps for completing this activity are listed above. The product, as mentioned, is the "program information requirements matrix."

| | Program Teachers | Program Director | Superintendent | School Board | Local Community | SEA | USOE | Professional Community |
|---|---|---|---|---|---|---|---|---|
| Basic Skills Development | ++ | ++ | ++ | ++ | ++ | ++ | ++ | ++ |
| Career Education Learner Outcomes | ++ | ++ | ++ | ++ | ++ | ++ | ++ | ++ |
| Program Costs | o | ++ | ++ | ++ | ++ | ++ | ++ | ++ |
| Teacher Attitudes | ++ | + | - | - | - | o | - | + |
| Teacher Behaviors | ++ | + | + | - | - | o | - | + |
| Use of Classroom Time | ++ | + | + | - | - | o | - | + |
| Use of Instructional Materials | ++ | o | - | - | - | o | - | + |
| Overall Program Implementation | - | ++ | ++ | - | - | + | - | + |
| Program Management | - | ++ | ++ | - | - | + | - | + |
| Community Acceptance | o | ++ | ++ | ++ | - | ++ | ++ | + |
| Parent Involvement | + | ++ | o | + | - | o | - | + |
| Use of Para-professionals | + | + | o | - | - | o | - | + |

FIGURE 2

Example of a Completed Domain Analysis Matrix for a Career Education Program

++ Essential Information
+ Important Information
o Marginal Information
- Information of Questionable Utility

| | Program Teachers | Program Director | Superintendent | School Board | Local Community | SEA | USOE | Professional Community |
|---|---|---|---|---|---|---|---|---|
| **First Priority** | | | | | | | | |
| Student Cognitive Growth | | | | | | | | |
| Student Affective Growth | | | | | | | | |
| Program Costs | | | | | | | | |
| Community Acceptance | | | | | | | | |
| **Second Priority** | | | | | | | | |
| Parent Involvement | | | | | | | | |
| Teacher Behaviors | | | | | | | | |
| Use of Classroom Time | | | | | | | | |

FIGURE 3

Example of a "Program Information Requirements Matrix"
Developed from the Domain Analysis Matrix in Figure 2.

| | Program Teachers | Program Director | Superintendent | School Board | Local Community | SEA | USOE | Profession Community |
|---|---|---|---|---|---|---|---|---|
| **First Priority** | | | | | | | | |
| Student Cognitive Growth | | | | | | | | |
| Student Affective Growth | | | | | | | | |
| Program Costs | | | | | | | | |
| Community Acceptance | | | | | | | | |
| **Second Priority** | | | | | | | | |
| Parent Involvement | | | | | | | | |
| Teacher Behaviors | | | | | | | | |
| Use of Classroom Time | | | | | | | | |

**FIGURE 3**

Example of a "Program Information Requirements Matrix"
Developed from the Domain Analysis Matrix in Figure 2.

# 3

## IDENTIFY CONCEPTUAL FRAMEWORK FOR EVALUATION

The conceptual framework is the philosophy or approach which guides the evaluation effort. Numerous models have been developed in recent years which provide a theme for focusing the evaluation process. The Information Based Evaluation (IBE) approach recommended for your use by National Testing Service is one such model. However, recognizing that there are several models from which to choose, it is important that you select that approach which most likely will satisfy your unique needs. The majority of the evaluation models appearing in the literature have several commonalities; however, each has a unique contribution and emphasis.

Of the several models available, some are more appropriate for different situations and programs than others; thus, a thorough review of the options is useful. One quite useful approach in selecting an appropriate conceptual framework is to synthesize the most applicable aspects of several models with the evaluation team's own experiences. This approach is much more desirable than the often encountered situation of adhering blindly to a particular model just because it has been widely discussed in the

literature; utility should always be the primary determinant in selecting a conceptual approach.

Several suggested process steps for completing this activity are listed below:

a. Review available evaluation models.

b. Review information requirements.

c. Synthesize a summary of the purpose and intent of the evaluation.

d. Summarize program information mandates.

The product for this activity is a concise statement of the conceptual framework which will be applied to the evaluation effort.

# 4

## SPECIFY EVALUATION QUESTIONS OF INTEREST

The completion of this activity by those in attendance at the design conference will determine the quality of your evaluation. The product of this activity will be the formulation of evaluation questions of interest. The importance of a well defined list of evaluation questions cannot be underestimated.

The development of evaluation questions, more than any other activity, dictates the scope and quality of the evaluation effort which is to follow. It is incumbent then, on all those in attendance at the design conference, to attend to the proper formulation of questions. The answers to the questions will provide the information that decision makers need to make decisions about the career education program in their district.

A common shortcoming made in evaluating educational programs is to build or design the evaluation around a set of program objectives. Since most programmatic outcomes are stated in terms of measurable objectives, this is understandable. However, such a focus provides an evaluation that is inadequate for at least two reasons: (1) the analysis techiques developed in the evaluation

a non-career educational program?" (2) "What types of students gain more career awareness from the program, and what types of students gain least?" (3) "Are major process objectives being attained on time?" (4) "How are teacher attitudes towards career education related to program implementation?" These are but a few examples that can be listed of the types of questions that should be generated in this activity. The reader should note that the questions are clear, concise, and in conversational English.

The range of evaluation questions that should be developed is defined by the outcome of Activity 2--the Program Information Requirements Matrix. The atmosphere at this phase of the activity is one of a "think-tank," with all the participants in the design conference formulating questions which they, as decision makers for the various facets of the program, have an interest in seeing answered.

In summation then, the sequence of activities for Activity 4 is as follows:

1) In this activity we review the program information matrix as developed as the outcome of activity.

2) All in attendance at the design conference should conceptualize evaluation questions for each cell in the informational requirements matrix as developed in Activity 2.

3) An interactive process of negotiation by all in attendance at the design conference should develop a tentative list of questions.

4) Procedures should be developed at this point which allow for the addition of new evaluation questions or developed evaluation questions as the evaluation proceeds and as new information requirements become evident.

The product of Activity 4 will be a list of primary and secondary evaluation questions of interest.

should relate to a broad spectrum of information requirements and management needs and not to the program objectives; and (2) objectives should not be treated as the foundation for the evaluation, because this limits the possibility for providing information for relating one objective to another.

The foregoing is not to be taken as a denouncement of evaluation by objectives. It does, however, point out that designing an evaluation strictly on measureable objectives limits the scope and therefore the utility of that evaluation at a later stage. The starting point for developing evaluation questions of utility should be the cell or a number of cells identified in Activity 2. From this, specific sets of evaluation questions can be developed for each cell of the Program Information Requirements Matrix. Each set of evaluation questions, therefore, addresses a specified domain of information as it pertains to a given class of information user.

The questions that are developed should be stated in conversational English. There are three reasons for this preference for a conversational format: (1) to facilitate the understanding of the process by non-evaluation oriented personnel at the design conference; (2) to emphasize the fact that all evaluation questions cannot be answered using traditional quantitative analysis techniques; and (3) it is much easier to get consensus on an evaluation need that everyone understands.

Several examples of evaluation questions related to career education are: (1) "Do secondary students in the career education program gain more on achievement tests than similar students in

# 5

## SPECIFY DATA BASE STRUCTURE AND CONTENTS

The outcome of Activity 4 will allow you to construct a data base to be used for the evaluation. The data base is the central source of feedback and information for the evaluation and the project. The structure of the data base will be dependent upon the number of data elements involved. The number of data elements involved will be dependent upon the number and type of evaluation questions that have been developed in Activity 4.

A technique which has been developed to identify required data elements is question mapping. The question mapping technique involves breaking the evaluation questions into key words and phrases and then semantically expanding or defining these terms. Through mapping only those data elements which are required to answer evaluation questions, a saving of time and money will result which might otherwise have been spent to collect superfluous data.

An example of the mapping technique as applied to the hypothetical questions developed in Activity 4 illustrates its utility.

> Do new students in the career education program (treatment students) gain more raw score points (statistical and practical significance) on an achievement battery (measurement instrument) than do like students (control group) in a more traditional classroom?

By applying the mapping technique, the data elements are identified which are needed to answer the evaluation questions. These data elements are treatment students, raw score, statistical and practical significance, instrumentation, and control group. The structure of the data base is determined by assessing the number of data elements, the resources available, and the type of manipulations that are required to answer the evaluation questions. The options are: mechanical, manual, and electronic. For most educational evaluations the computer is the most cost-effective approach.

After the data elements are identified, they are arranged and filed according to some previously developed classification scheme. The most common way to classify data elements is by the audience. The data file is a list of data elements; examples of five abbreviated files (student, teacher, counseling and guidance personnel, building, and process) are as follows:

Student File

1. System ID.
2. School ID.
3. Grade.
4. Student ID.
5. Name.
6. Race.
7. Sex.
8. Group: Project-Comparison.
9. Birth Day - Month - Year.
10. Basic Skills - Reading
11. Basic Skills - Mathematics
12. Occupational Self-Awareness
13. Work Values
14. Interest in Career Preparation
15. Attitude Toward Work

16. Knowledge of Work Values.
17. Decision Making.
18. Job Seeking/Job Getting.
19. Career Awareness.
20. Economic Awareness.
21. Academic Perceptions.
22. Social/Personal Perceptions.
23. Exposure - Items 1-27, Part II, Student Information Form.

Teacher File

1. System ID.
2. School ID.
3. Teacher ID.
4. Age Range.
5. Highest Degree Held.
6. Certification.
7. Employment Status.
8. Attitude Toward Career Education.
9. Knowledge of Career Education.
10. Skills in Career Education.
11. Grade(s) or Subject(s) Taught.
12. # Years Taught - Total.
13. # Years Taught Continuously in This School.
14. # Continuous Months Taught in Current Area.
15. # Classes Taught per Day.
16. # Students per Class.
17. Process.
    - Use of Materials.
    - Staff Development.
    - Instructional Strategies Used.

Counseling and Guidance Personnel File

1. System ID.
2. School ID.
3. Counselor ID.
4. Age Range.
5. Highest Degree Held.
6. Certification.
7. Employment Status: Permanent, Temporary, Prob.
8. Attitude Toward Career Education.
9. Knowledge of Career Education.
10. Skills in Career Education.
11. # Years Taught.
12. # Years in Counseling and Guidance.

## Building Data File

1. System ID.
2. School ID.
3. Lowest Grade Housed.
4. Highest Grade Housed.
5. # Administrative Staff.
6. # Counseling and Guidance Personnel.
7. # Specialists.
8. # Paraprofessionals.
9. # Volunteers.
10. # Students in Each Grade.
11. # Teachers in Each Grade.
12. Socioeconomic Level of Student Population.
13. White Student Population - %.
14. Mobility Rate - %.
15. Dropout Rate - %.
16. Average Tenure of Teachers.
17. Teacher Turnover Rate - %.
18. Federal Porjects in Building.
19. # Books in Library.
20. Teachers Teaching out of Certificate Area - %.
21. Teachers Participating in Staff Development - %.
22. Classroom Structure.
23. Length of School Day.
24. Process.

## Process File

1. Number of Staff Involved.
2. Number of Administrative Staff Involved.
3. Amount of Funds Expended.
4. Amount of Funds Proposed.
5. Program Implementation Index.
6. Time Spent in Treatment.
7. Types of Materials Used.
8. Use of Paraprofessionals.
9. Use of Volunteers.
10. Number of Field Trips.
11. Number of Out of School Experiences.
12. Number of Shadowing Experiences.
13. Number of Students Involved by Grade and School.
14. Number of Students Involved by Sex.

The length of each file, whether it be student, staff, or any other, will be limited by the number and type of questions determined by Activities 2 and 4. A word of caution should be extended at this time; if certain developed evaluation questions are not addressed in this as part of a data base, a rationale for that should be documented.

In summation, several suggested process steps for completing Activity 5 are as follows:

1. For each evaluation question of interest, four areas or phrases should be underlined: the actions, the audiences which will be either recipients or observers of the action, the outcomes, and the inputs.

2. For each phrase, list elements or characteristics which will define that phrase.

3. Eliminate redundant data elements.

4. Create data files by listing data elements by audiences, participants, and actions processes.

5. Select a data base structure-will it be electronic, mechanical, or manual? (The product of the foregoing steps will be a file(s) of data elements and a structure for the file.)

# 6

## SELECT INSTRUMENTATION

When the word "instrument" is used in education or other fields, an image of a carefully designed, pretested, validated questionnaire or test usually comes to mind. However, an evaluation instrument can be any procedure by which information is gathered and recorded; as such, "instrument" can take many different forms. Although the questionnaire or the achievement test are indeed frequently used instruments, so are notes of discussions, records of observations, and monitoring checklists. The instrument's form is secondary; it is merely the expression of the type of information desired, the sources of that information, how great a quantity of information is required, and so forth.

Probably the most important point that can be made about selecting an instrument is to first carefully state the questions for which you need information; do not select the instrument first and then try to make your questions fit. You will get answers to those questions, but those answers probably will not meet the information needs of the project. As in most situations in educational evaluation, an important caveat must be added: occasionally you may have little choice in instrumentation; legal mandates and/or practi-

cality may dictate. When this occurs you should make special efforts to ensure that those involved with administration of the career education program are fully aware of the information "costs" that will result.

The second most important point can be stated in various ways; but, all the ways boil down to one common prescription: all other things being equal, the best instrument is the one that provides the most useful information at the least cost. As used here, "cost" includes a variety of different components, and "least cost" inherently requires a consideration of trade-offs.

Some of the costs involved include actual monetary outlays for commercially prepared instruments, the time needed for development of your own instruments, difficulties in analyzing and reporting results, the possible need for trained administrators, and the burden on respondents. Some of these costs can be readily computed. For example, if a commercially available instrument can be located that closely matches your information needs, the publisher should be able to tell you the costs involved in obtaining, administering, scoring, and reporting. Frequently, however, the publisher will not be able to tell you of some other costs. Such unknowns might include how much time it will take your respondents, whether the items themselves are at a suitable reading level or use the "language" of your program, whether there may be delays in scoring results, or whether previous applications of the instrument have been normed on sufficiently similar populations to justify its use with yours.

Given these warnings, the suggested process steps for selecting the appropriate instrumentation are relatively straightforward.

Following a discussion of these steps, with examples describing the instrument selection process used in this present evaluation effort, is a brief presentation on each of the instruments NTS has utilized in this effort.

Steps in Selecting Instrumentation

1. List most appropriate sources of information needs based on previous activities. In the NTS evaluation design, the following were listed as sources of information:

   - Students - These students were generally in grades 5, 8, and 11, though these grades varied from place to place. The grade levels were chosen to reflect what OCE terms the "natural breakpoints" in career education programs and to allow fairly straightforward longitudinal comparisons at the conclusion of an additional three years. Further, a deliberate effort was made to select groups of students determined a priori likely to vary in their levels of exposure to career education. It was necessary to include students as a data element to measure learner outcomes, gain information about the effects of social and personal characteristics, and gain self-reports of exposure to career education activities. That is, students were selected as sources of input and process domain information and as the primary source of outcome or product data.
   - Teachers, Counselors, and Principals - The NTS design was based on the assumption that student outcomes

would be proportional to the extent to which the students had been exposed to career education related activities. Although students were asked about experience with career education activities, NTS felt that many would not be able to provide a very complete picture; further, this self-report problem would likely become greater the more completely the activities were infused into the curriculum. Teachers, counselors, and principals were seen, therefore, as additional appropriate sources for this process domain information. Counselors and principals were further seen as the best sources for their own particular activities. Teachers, and certainly students, were not expected to be able to adequately reflect those activities in their self-reports.

It was also hypothesized that input characteristics of the professional staff and of the school itself were likely to have an effect on learner outcomes. Teachers, counselors, and principals were judged to be the best sources of their own characteristics with the principal being best situated to respond also to queries about the school.

- Career Education Project Staff - Since the preliminary site selection activities by NTS had indicated wide variability in terms of school system characteristics, community characteristics, and program emphases, an information need existed to document those differences.

If the evaluation had concentrated on one project, this added effort would not have been necessary; though, it should be noted, a single-project design comparing results in different schools may require a gathering of descriptive information about those schools.

- Documentary Evidence and Census Reports - Although project staff were likely to be good sources of the background information discussed above, it was expected that the burden placed on them would be seen as excessive if they were asked for all of it. Such documents as proposals, evaluations, interim reports, and final reports were judged to be acceptable, unobtrusive sources of program, school system, and community background data. Additional community information could be obtained through census materials; this source has the further advantage of presenting data collected using uniform procedures and definitions.

2. Review available data collection instruments to determine how well each matches elements in data files.

Three basic types of instruments are available for your use: (1) those you generate yourselves, (2) those that are available commercially from publishers, and (3) those that may have been non-commercially developed to fit the needs of projects similar to yours. Which type you select should be based on how well an instrument will answer your questions for the least cost. No hard and fast rules can be laid down to determine your choice. Each type has its advantages and

disadvantages; again, you must judge the trade-offs. One of the most important tradeoffs will depend on the type of design you select. Following the general discussion of these different instrument types, problems of utilization in longitudinal efforts will be discussed.

Commercially Available Instruments

Publishers of test materials tend to concentrate on "product" or "outcome" areas. You are undoubtedly familiar with a number of achievement test batteries, self-concept measures, and specific outcome instruments such as the PECE, Ohio, and New Mexico series used in the first year of this evaluation.

- Advantages
  - Ready availability. To be sure, most publishers will bring them to your attention.
  - Generally pretested. Most publishers will not offer instruments that have not been developed by those skilled in the craft, that have not been piloted, and on which there is not at least some data concerning the reliability and validity of the results obtained.
  - Ease of scoring the results. Although the purchase price of many instruments, particularly those designed for administration to students, will include the cost of scoring and reporting the results, others allow you the option of scoring the instruments yourself and provide you the necessary guidelines for doing so.

- Disadvantages
  - High initial cost. A price of a dollar or more per student is not unusual. It is important to remember, however, that this high initial outlay may mean savings to you and your staff later on.
  - General focus of instruments. In order to be available to the widest possible market, most commercial instruments will purport to measure a wide variety of learner outcomes. If your questions are narrower, you will be paying for non-essential information.
  - Narrow focus of instruments. This is the other side of the coin. Although many instruments purport to cover a broad range of outcomes, the results of NTS's instrument substudy indicate that this may not always be the case. If you have chosen to use commercially available instruments, this may mean having to use more than one, thus increasing your costs and the burden on respondents.
  - Appropriate to your respondents. Most commercial instruments are designed to fit the average respondent in a group, the average fifth grader, for example. Your fifth graders may be, as a whole, above or below this average. The result can be frustration either from boredom or inability to understand. Results obtained from frustrated fifth graders may not be terribly valid.

Given these advantages and disadvantages, information about the specific commercial instruments used in the national evaluation can be found in the Evaluation Design Document and the appendices to the Final Report prepared by NTS.

Further material is available in A Functional Guide to Evaluating Career Education (prepared by Development Associates, Inc. for DHEW, in September 1975), the compendia of materials being developed by the National Center for Career Education of the University of Montana, the Career Education "Whole Earth Catalog" currently being produced by Pacific Consultants of Washington, D. C., and the Report on Career Education Evaluation efforts completed by New Educational Directions, Inc., of Crawfordsville, Indiana.

For a general exploration of NTS's instrument selection process, please see Chapter 6 of the Evaluation Design Document. For a complete analysis of the instruments selected, see the Career Education Instrumentation Substudy, Volume III, of the Final Report.

If you choose to continue the implementation of the longitudinal design, then the instruments you select should be those you found generally satisfactory and which addressed the questions of importance to you used in the first year. The Final Reports and the Instrumentation Substudies should help you make a decision on which instruments to continue to use.

Already Developed Non-Commercial Instruments

Frequently, you already may have some of the instruments you need. Perhaps they were developed in previous years by

project staff, outside consultants, or state agency personnel. It is NTS's experience that these instruments tend to focus on formative concerns rather than learner outcomes. If the former are important information needs, then these instruments can prove very useful.

- Advantages
  - Inexpensive. As these are non-copyrighted, the only direct costs may be for duplication. Further, little additional staff or consultant time, beyond locating and "customizing" them is required for developmental purposes.
  - Appropriate. Usually such instruments have been developed to meet the information needs of projects like yours.
  - Ease of scoring. Usually these instruments are very brief, making hand scoring a pragmatic option.
- Disadvantages
  - Possibly non-reliable and/or invalid. That an instrument has been developed to gain a certain type of information in no way guarantees that it actually does so. This is particularly the case when attitudes or knowledge are being accessed. "Face validity" is only one indicator of appropriateness; one of the unfortunate facts of evaluation is that our respondents frequently do not react to questions the same way the evaluators do. This can lead to an additional set of problems involving interpretation of the results.

- Interpretability. Locally produced instruments may be prone to one or another (or even both) of two distinct problems: (1) They may include highly ambiguous questions and provide, therefore, less than useful answers and/or (2) they may be replete with "motherhood" questions. Answers to the latter may serve to impress, but their contributions to analysis are minimal.

In short, if instruments of this type are available, meet your needs, and you are willing to take the risks indicated above, using already developed, non-commercial instruments can be a very cost-effective approach to evaluation.

Self-Generated Instruments

It is likely that your reviews of available instrumentation (commercial or other) will not uncover any that fit your needs or, at least, fit all of them; or, you may have found what you needed, but the costs to your project may be too great. If any of these occur, you are left with three choices: (1) abandon the effort to collect information in a particular area (which should only be done following full and open discussion with project staff); (2) modify an existing instrument to make it appropriate (which is no problem, usually, with non-commercial instruments but should be done gingerly with commercial ones given copyright constraints); or (3) develop your own. Some guidelines for doing the latter are described below. These guidelines will also apply to the second option; the first is a matter for you and the program staff.

- Advantages
  - Appropriateness. Given your knowledge of your own information needs, the instrument can be designed to match them with a precision unobtainable elsewhere. Further, you can design them to fit the abilities (such as reading levels) of your respondent groups.
  - Efficiency. It may be that developing and producing your own instruments may be faster than depending on others. Further, you are much less likely to be in the position of scoring extraneous responses if you have limited the questions to only those that are essential.
- Disadvantages
  - Cost. Frequently, developing your instruments may actually cost more in staff time than purchasing them elsewhere. This is especially the case when the information needed requires a detailed questionnaire or survey.
  - Other disadvantages. The disadvantages listed under the other two types are equally relevant here.

Suggestions for Developing Your Own

If, after reviewing the other possibilities, you find that you must develop your own instruments, there are a few general rules which we at NTS have found to be useful.

- Readability. If you are preparing questionnaires or other survey instruments, it is important to insure that the reading level is appropriate to your respondents. This is an obvious point when younger students are the respondent group but is equally important when asking questions of others. The use of jargon or abbreviations which may be a part of your intensive involvement with career education may only draw a blank with others. It is always advisable to pilot your instrument on a few people before going into production.
- "Answerability." Questions phrased from an unfamiliar perspective can lead to respondents not answering or answering randomly. Frequently, this can lead respondents to treat the whole effort in a cavalier manner. High school students, in particular, may end up making airplanes rather than providing you with answers.
- Variance. If your evaluation is concerned only with discovering the answers to descriptive questions, consensus in responses will not be too harmful. At the same time, if you desire to understand the relationships between (among) two (or more) items of information, then your questions should generate some disagreement across the responses.

- Length. An endemic problem for instrument designers is knowing when to stop. You must be sensitive to the fact that some questions are more important than others, and that respondents do not appreciate an excessive burden.
- Scoreability. When designing your instruments you should constantly keep problems of data reduction in the forefront. If your results are to be keypunched, you can prevent later aggravation by conferring with keypunchers about preferred formats. Similarly, if hand scoring is to be used, keep ease of tabulation in mind.

NTS found it necessary to develop a number of instruments for this evaluation because none existed that were appropriate to the questions being asked and/or for the people of whom they were to be asked. In fact, only in the outcome domain were project staff able to locate suitable instruments. New instruments had to be developed for the context, input, and process domains. The instruments developed are located in the appendices of the common Final Report. Analyses of the instrumentation are included in Volume III of the Final Report, as well.

Two further general points need to be made about instruments selection. First, if suitable information is already being collected (such as achievement test results for students), then it is far more efficient to adapt its use than to attempt to duplicate it. Second, all instrument selection is going to require tradeoffs and compromises; keep the program staff informed.

## Instruments in the Longitudinal Design

Simply stated, if you wish to gain comparable information over time which is, after all, the rationale for longitudinal designs, then you should use the same instruments to collect it. This does not mean that you must use the entire battery of instruments utilized during the 1976-77 project year, but, depending on your information needs, you should use those instruments that fit those needs. The learner outcome instrumentation substudy results, included as Volume III of the common Final Report, can assist you in making these choices.

If you adopt a panel design (see Activity 7), then an additional problem of instrument selection arises. That is, which level of the instrument should be used at the various grade levels? NTS's validation efforts concentrated on the fifth, eighth, and eleventh grade levels. To test at other grade levels, NTS suggests using the "one up, one down" rule of thumb. In other words, if an instrument is valid at the fifth grade level, it is likely not to be too invalid at the fourth and sixth.

# 7

## SPECIFY POPULATION SELECTION PROCEDURES

Given the variability of sample size and characteristics across the fifteen projects in the NTS national evaluation effort, it is virtually impossible to detail the steps each project should follow to continue the longitudinal study. For example, the project samples involved ranged from 150 students to over 3000; the grade levels covered included third, fifth, sixth, eighth, ninth, tenth, eleventh, and twelfth. The number of schools ranged from one to nearly thirty in any one site at any one grade level. This does not mean that no suggestions can be offered; rather, the suggested steps may require some modification to fit your project's first year sample characteristics.

The modal pattern of first year samples included several hundred students at each of the fifth, eighth, and eleventh grade levels; longitudinal sampling considerations for this group will be the general framework for this discussion. The sizes of your samples are included in the Local Report prepared for your project.

The samples were selected in most cases to provide sufficiently large numbers to allow continued surveys of the same respondents

throughout the remaining years of the design. That is, generally enough students were selected for use in a panel type longitudinal design. The statistical justification for the original selection process can be found in Chapter 7 of NTS's Evaluation Design Document.

Since your initial sample has been selected already, this section will concentrate only on those sampling concerns directly related to the longitudinal design. Briefly stated, these concerns include the use of a panel design and/or the use of a "replicability" design, steps for following-up or tracing the students and others in the original sample, an expansion of the size and/or scope of the sample, a comparison of the input characteristics of the samples in successive years, and the selection of control groups.

## Types of Longitudinal Designs

Two different types of longitudinal designs are available for your use. Your selection should be based on the nature of the evaluation questions worked out with the program staff and the characteristics, especially size and grade level, of those in your original sample. These types are panel designs in which the same individuals are surveyed throughout the span of your evaluation and replicability designs in which different groups of individuals sharing similar characteristics, such as school and grade level, are tested in successive years. The choice of model depends, again, on which questions you desire answered and the characteristics of your original sample.

A panel design's primary advantage lies in asking the same (or many of the same) questions of the same individuals across

time. Much as a one year pre-post design permits inference about gains, a panel design allows these inferences to be made across successive years. In the ideal situation, your career education evaluation effort might involve the survey of the same students for three years with each year including a pre and post test. This would allow you to make yearly inferences concerning student gains and, more importantly, to make statements of cumulative gains as well.

There are also some significant disadvantages to this design, however; these include the need to account for problems involved in sample attrition and the high costs of following individual students. If your information needs require continued process domain "testing" of professional staff, you also will need to make detailed selection plans for surveying the appropriate personnel. Some suggestions for avoiding or, at least partially, ameliorating these difficulties are discussed in later segments of this section.

Replicability designs, on the other hand, permit only the most general statements about cumulative program effects though, depending on your information needs, these may be sufficient. If, for example, the 1976-1977 school year was the first year of full K-12 implementation of career education in your school system, repeated testing of each school year's fifth grades will provide some information on the cumulative effects of earlier elementary school efforts. This method's main advantages are in a more practical vein, however. It is usually easier to survey the same classrooms in successive years than it is to trace individuals; along the same lines, it also would probably be easier to institutionalize such a method as part of your system's annual assessment efforts. A further advantage

is that you can avoid problems stemming from panel attrition especially if your original sample was fairly small.

Another set of advantages for a replicability design involves instrumentation. Instruments found to be relatively valid and reliable at a given grade level can be used with confidence at the same grade level again, and the possibility that students are "learning the test" is avoided. However, there is still a possibility that teachers whose students are repeatedly tested may begin "teaching to the test." For additional information on problems of instrumentation in longitudinal evaluation, see Activity 6.

Two major disadvantages are involved with this approach. The first is that the permissible evaluation questions are limited only to group comparisons across time. Statements about individual gains are restricted to those possible within the pre/post period. The second is the need to pay careful attention to the characteristics of the groups in different years. As the analysis of input characteristics shows (see the Final Report), career education learner outcome differences are partially a result of such input differences.

It should be noted that these two methods are not mutually exclusive. For your purposes, it may be necessary to use both. For example, you may wish to follow-up the eighth grade panel while, at the same time, replicating the fifth grade study.

## Following-Up Your Respondents

If you choose to use a panel design, then, careful tracing of the students involved in the first year is mandatory. This segment contains some suggestions for doing this and discusses some of the problems you may encounter in the effort.

Based on an expected annual attrition rate of each of the student samples of 20 percent (the approximate national mobility rate), it is clear that your sample sizes will be much smaller at the end of the third year than they are now. The mobility rates for your school system are included in the Evaluation Report prepared for your project by NTS. It should be noted that this rate can vary markedly from year to year as a result of many external economic or social factors.

If your samples were relatively small to begin with then, clearly, you will have to make special efforts not to "lose" students who have not actually left the system. Even if your original samples were quite large, careful tracing may be necessary to prevent building in additional biases. This could occur, for example, if some types of students move within the system more frequently than other types.

NTS has supplied you with the following information about each of your students to assist your follow-up activities:

- ID Number - In many cases this ID number is the same one normally used in your system for the particular student. In other cases, the ID numbers can be matched to individual student names on the original pretest packet envelopes.
- Teacher's ID Number - If the students were sampled on the basis of classroom, this information will help you to assign student ID numbers to the students.

- School and Grade - As with the teacher ID number, this information will assist you in a similar manner.

This follow-up task, already complicated by sample attrition, is made even more difficult by the classroom "cluster sampling" technique used by NTS in the first year and by the selection of grades 5, 8, and 11. In most systems, students in these grades will almost always change classrooms from year to year and are also likely to be changing (or leaving) schools as well. If you choose to continue test administration to only those students in the original sample (see next topic for alternatives), you may need to formulate plans for not only tracing the individual students but also for some type of centralized testing situation.

Tracing teachers, counselors, and principals should be a much easier task. NTS has supplied you with a complete list of the ID numbers and schools for each individual in these groups as well as the grade level, where relevant. NTS expects only about a 5 percent annual attrition rate among these personnel based on national averages; of course, external factors such as budgetary constraints may make your rate somewhat higher. If you choose only to use the students sampled during the first year, then, of course, the teachers (and possibly the counselors and principals) surveyed then may not be appropriate data elements in subsequent years. In this situation the relevant professionals would be at the 6th, 9th, and 12th grade levels for the second year, and the 7th and 10th for the third.

## Expanding the Size of Your Sample

Two important reasons exist for considering the expansion of the sizes of your samples beyond the sizes of those used in the first year: (1) your own evaluation questions may require information from more students and/or a broader range of grade levels of students; (2) the pragmatic difficulties involved in tracing individual students and providing centralized testing for them (and, thus, requiring classroom disruption) may make it expedient to continue sampling by classroom.

If you choose to expand your sample in order to answer your own evaluation questions, Chapter 7 in the Evaluation Design Document and the sources cited there will provide guidance for selection procedures and numbers required.

If you determine that your evaluation questions can be addressed through the panel alone but also judge the costs to be too high for such selective testing, the following few general points may assist you. NTS has found the classroom cluster sampling technique to be not only cost-effective for actual test administration but also to cause less disruption within the school setting than actually removing the individual sample students from their classrooms. It should be noted as well that classroom sampling may help obviate attrition-based problems of sample size.

Nonetheless, such classroom sampling within a panel longitudinal design raises a number of problems. Among these are comparability of the samples across time, similarity of the original sample students to those who are added, and either a risk of

major loss from attrition or a drastic increase in sample size. These problems are especially likely to arise when there is a great amount of pupil reassignment from year to year. In such cases, careful comparision of input domain characteristics will be necessary to trace differences between years and between original panel students and the others. The following topic and the discussions and analyses of input characteristics in the Career Education Final Report will assist you in this area.

## Comparing Input Characteristics

NTS devoted much attention during this project year to evaluating the effect of input characteristics on learner outcomes. For a review of these findings, please see the Final Report for the national study and its appendices, your project's Final Report, and the Evaluation Design Document.

It is clear from those findings that it is important to know a great deal of information about the respondents. This need is particularly acute in longitudinal evaluations whether the design used is based on the replicability method or the panel approach, even with an expanded sample. If a replicability design has been chosen, for example, then to make any inferences about yearly differences, it will be necessary to examine and then control for differences in the input characteristics of the group.

The same is true of panel designs. To be sure, unless the composition of the panel is monitored carefully, you may end up making some incorrect inferences about program success. Some

attrition at higher grade levels results from less skilled or less motivated students dropping out of school. Thus, even if individual students were only maintaining the same learner levels, the average of the sample would probably show an increase.

## Control Groups

Based on the long-range plans summarized in the FY 1977 proposals from the projects involved in this evaluation, there should be few, if any, control group students available in the next few years. Nonetheless, in order to address your evaluation questions, it may be necessary to compare students with high exposure to career education concepts and activities to students with low exposure.

In such cases, the determination of expected exposure levels is properly a concern of the program staff. NTS has found, for example, that "uncontaminated" students are rare indeed. Following the identification of such students, however, it is the evaluator's responsibility to ensure as much input characteristic similarity as possible in those variables identified in the Final Report as being important. The actual numbers necessary and other selection requirements are covered in Chapter 7 of the Evaluation Design Document.

# 8

## SPECIFY DATA ANALYSIS TECHNIQUES

With knowledge of the nature of the evaluation questions, the data elements, the instruments, and the population selection procedures, the appropriate analysis techniques can be selected. However, it is frequently the case that only the statistical expert has the knowledge and experience necessary to make the judgment as to what would constitute an appropriate technique. The intent of this section is not to present a "mini-cookbook" on statistical techniques, so that the advice of a statistician need not be sought. Rather, the intent of this section is to present for you a useful framework for conceptualizing some common career education questions and some common analysis techniques for addressing those questions.

Career education evaluation questions can be usually grouped into question categories:

- Questions of description
- Questions of comparison
- Questions of relationship
- Questions of prediction.

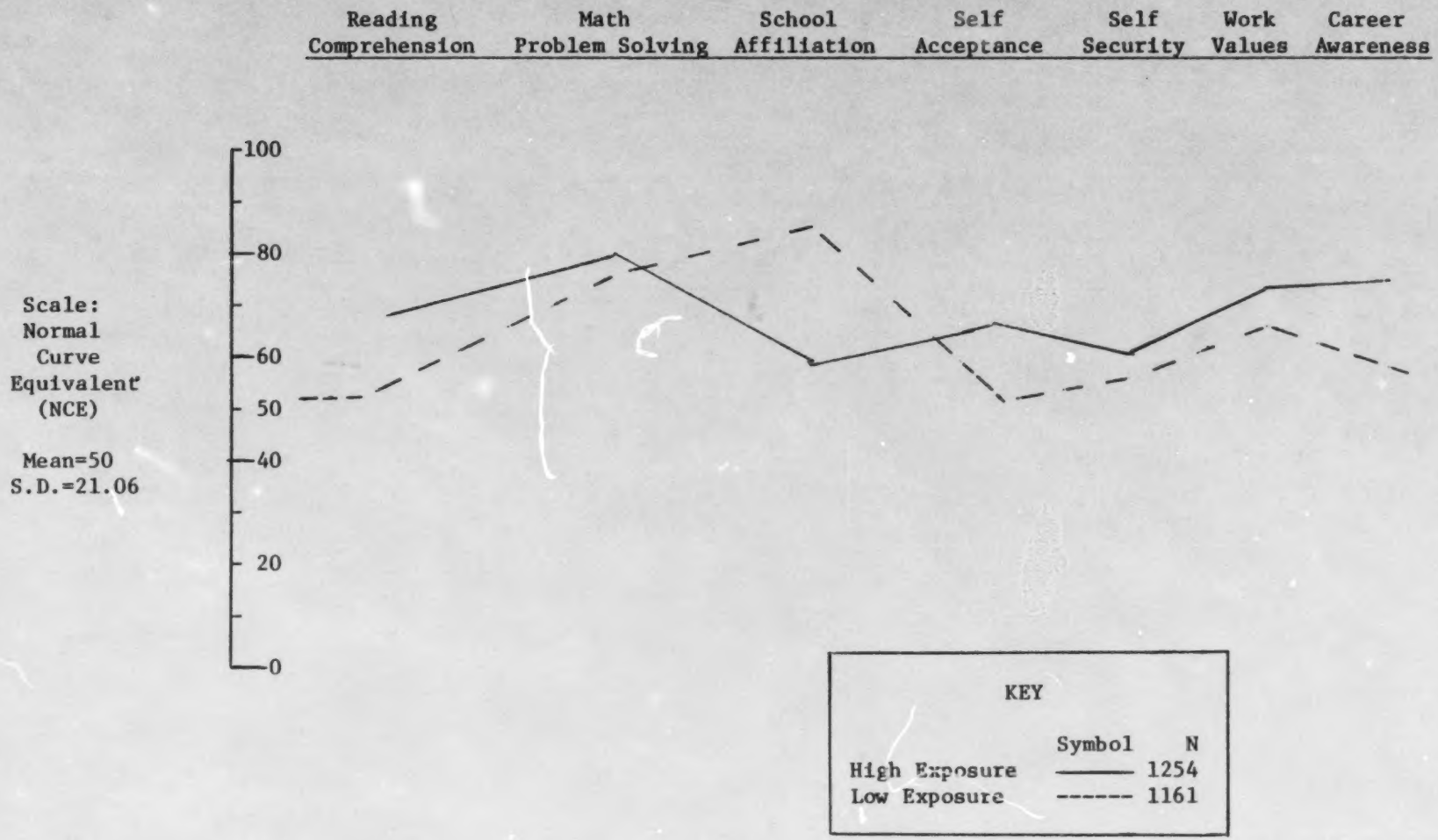
Questions of description are answered with simple descriptive statistics, such as standard score means, standard deviations, counts, and profiles. Graphic display is particularly applicable in answering this type of question. Questions of comparison deal with contrasting one or more groups of individuals, classrooms, or buildings on a set of input, process, and/or outcome variables. One of the most important issues addressed by questions of comparison involves estimating the achievement of participants in career education classrooms on the learner outcomes without the benefit of a career education program. Later in the chapter we will explore several approaches to be used in the present evaluation for estimating the effects of "no treatment." Questions of relationships deal with understanding the interrelationships among the various input, process, and outcome variables. Questions of prediction address three subsets of questions. The first subset relates to straightforward questions of prediction. The second concerns questions of causal dependency, and the third deals with questions having to do with the effect of the variable, "time."

## Questions of Description

A number of important questions can be addressed in an evaluation by graphically presenting descriptive data. Many of you are probably familiar with the variety of ways graphics can be used to simplify data presentation. The following constitutes two such examples.

Figure 4 presents hypothetical adjusted posttest outcome profiles (adjusted for pretest differences) for students with high exposure to career education concepts and students with low

Adjusted Posttest Outcome Profiles for Grade 5 Students
Receiving High and Low Exposure to Career Education




8.3

FIGURE 4

exposure. In this fictitious example high exposure students outperform low exposure students on most of the outcome constructs, with the notable exception of School Affiliation.

Figure 5 depicts the process variables which show the strongest relationships to improve basic skills. Students who are exposed to these fictitiously represented career education strategies demonstrate higher basic skills achievement (reading and math composite) than do students not exposed to these strategies.

Figures 4 and 5 illustrate two ways in which graphic display of descriptive data can be used to augment narrative discussion in an evaluation report.

## Questions of Comparison

Perhaps the most important question addressed in a career education evaluation design is whether students exposed to career education demonstrate improvement on a wide range of student outcome measures over and above how they would have performed had they not been exposed to career education. The question centers on obtaining an estimate of no treatment effect (i.e., what the students would have learned with no treatment). In short, how much more did students learn by participating in the project than they would have learned if there were no project?

Answering this question involves comparing students designated as career education students with some estimate of what these students would have achieved had they not participated in a career education program. Because of the policy implications inherent in an answer to this question, it is recommended that several approaches

Correlation Between Basic Skills Achievement and Selected Career Education Dimensions

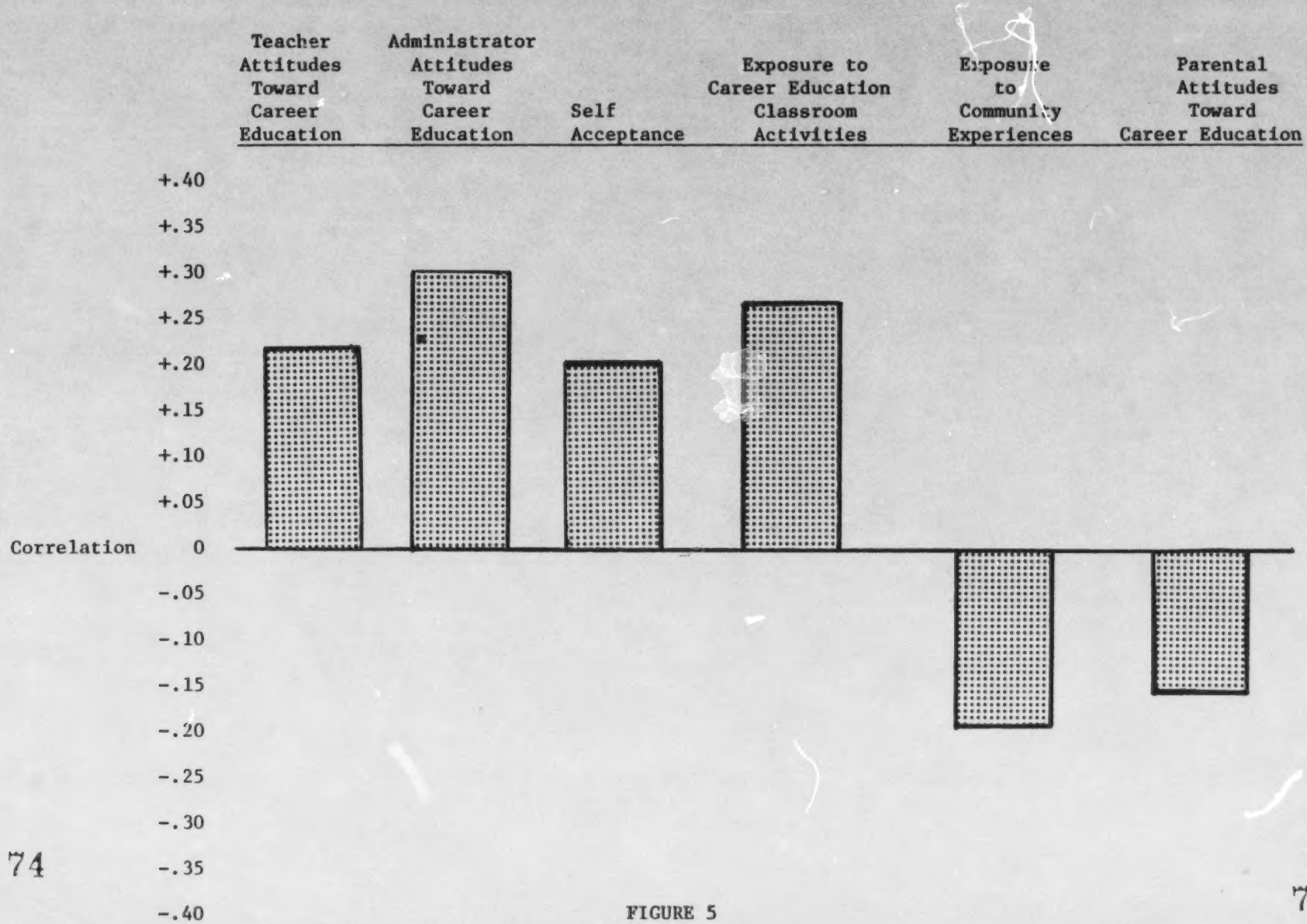



FIGURE 5

be used to estimate the somewhat elusive notion of no treatment effect. Specifically, it is suggested that three separate approaches can be employed simultaneously to estimate what students would have achieved had they not participated in the career education program. Each of these three estimates can then be compared with what career education students actually achieved. This will provide three estimates of effect. If all three estimates are not in the same direction, one would usually base conclusions on the "majority rule."

Norm-Referenced Approach

Career education students can be compared to a norm group comprised of a nationally representative sample of students at the same grade level. The no treatment expectation is that the project pupils will maintain, at posttesting, the same achievement status with respect to the norm group as they had at pretesting.[1] The strengths and weaknesses of this approach are summarized by Tallmadge and Horst as follows.

- Strengths. Where no comparison group is available, the norm group provides a plausible estimate of no treatment posttest scores. Even where a comparison group is available, unless it comes from the same population as the treatment group, the norm-referenced model offers a more defensible estimate of posttest performance at substantially less cost and effort than a comparison-group design.

- Weaknesses. The validity of the model rests on the assumption that the achievement status of a particular subgroup remains constant relative to the norm group over the pre- to posttest interval if no special treatment is provided. Empirical support for this assumption is minimal. It is

---

[1] Tallmadge, G. Kasten and Donald P. Horst. A Procedural Guide For Validating Achievement Gains in Educational Projects. Washington: U.S. Government Printing Office, 1976.

conceivable that some subgroups would move up and others move down in the normal course of events. When the norm group is like the treatment group, the plausibility of the underlying assumption is greatly enhanced; thus, for example, norms for gifted children would be best for assessing a project serving such pupils.

- Implementation Considerations. This model is widely applicable as it does not require a comparison group. The model requires the use of standardized tests. The same level of the same test should be used for both pre-and posttesting. Program participants may not be chosen on the basis of their pretest scores. Both pre and posttesting must be accomplished on dates corresponding to the ones on which the test publisher collected normative data.

The norm-referenced approach can be used in at least the basic skills and self concept domains because of the availability of national normative data. There is also the possibility that the data gathered for the NTS Common Evaluation will provide acceptable, albeit quasi, national norms for the career education instruments used in this study.

Comparison Group Approach

This approach involves pre- and posttesting both career education students and a similar comparison group which did not participate in career education. Analysis of covariance is used to adjust posttest scores for initial pretest differences before comparing the treatment and comparison group means. Tallmadge and Horst describe the strengths and weaknesses of this design:

- Strengths. When between-group differences are evident despite the random assignment, analysis of covariance provides the best method of adjusting observed posttest scores for random pretest differences. Comparing posttest means that have been adjusted is always more precise than comparing unadjusted posttest scores.

- Weaknesses. This model assumes that treatment and comparison students are random samples from a single population so that any difference in pretest performance is due only to sampling error and random error of measurement. It will not provide an appropriate adjustment for pretest score differences which reflect nonrandom differences between groups. Where analysis of covariance is employed with data from pre-existing, intact groups, there is always some danger in presuming that the groups are random samples from a single population.

- Implementation Considerations. This evaluation model allows a wide choice in the test instruments to be used and in the time of testing. If...a norm-referenced comparison is also made, the choices become more restricted. The control group must be very similar to the treatment group. True random selection is strongly advised. If the groups were not selected randomly, strong evidence is needed to demonstrate that the selection was "random in effect."[2]

The unique assumption of this approach is that the treatment and comparison groups are similar on all variables excepting participation in a career education program, and that any pretest performance differences are due to sampling error and random error of measurement. This requirement can only be met if students are truly randomly assigned to "target" and "non-target" groups. Fortunately there is a great deal of opinion that suggests this technique is of sufficient power to yield valid results even if students have not been randomly assigned --which is the predominant situation for career education evaluations.

Exposure Index Approach

The exposure index approach is based on the premise that if career education programs are having an impact over and above

---

[2]Tallmadge and Horst, pp. 49-50.

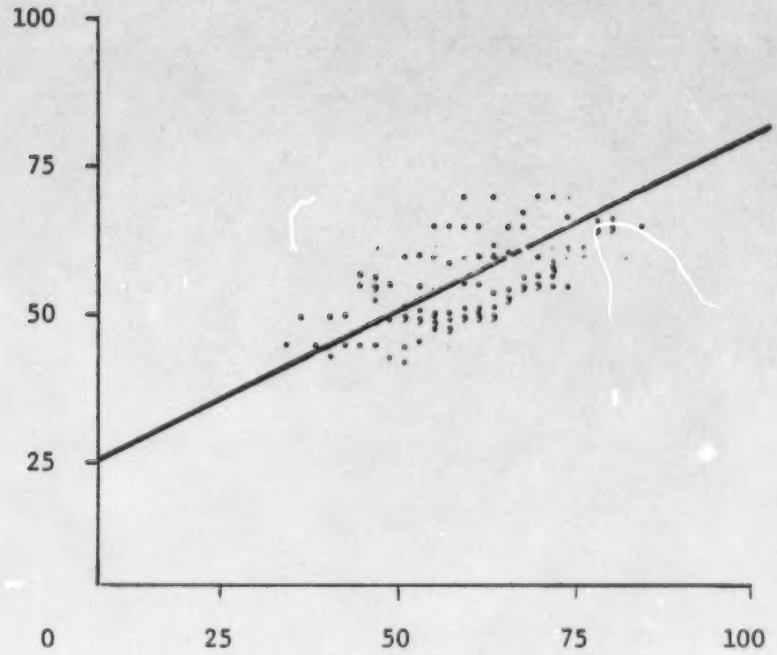
that which would be expected were there no program, then students exposed to higher amounts of career education concepts and activities should show higher outcome scores than students exposed to lower concentration. If career education outcome scores adjusted for pretest score differences are regressed on an exposure index, then the outcome value at the intercept of the regression line is the estimate of no treatment effect (see Figure 6).

The major assumptions of this approach are: (1) that there exists some set of strategies or activities that differentiate "career" education from "general" education; and (2) that the degree of exposure to these career education strategies or activities is positively and linearly related to improvement in career education outcome measures.

In summary, the above three approaches can be used to estimate no treatment effect in an effort to improve the validity of any conclusions regarding the impact of career education on student outcomes.

For those of you interested in going one step further in looking at the matter of no treatment effect, we suggest you use the joint effect model as explained in the Evaluation Design Summary chapter of Volume II of the Career Education Final Report. This particular model consolidates the last two approaches mentioned above into one analysis. In this model, both group membership and exposure are correlated with outcome score to the extent that the unique contribution of group membership to explaining outcome score variation (after entering pretest) is large. You can infer that there was some process effect on the outcome, but your exposure variables failed to specify the relevant aspects

Exposure Index Scores Regressed on
Adjusted Career Awareness Posttest Scores




CAREER AWARENESS SCORES

EXPOSURE INDEX

FIGURE 6

of the process. If, on the other hand, the unique contribution of exposure variables to explaining adjusted outcome variation is large, you are left with the conclusion that there was little difference between the treatment (program) and the nontreatment group with respect to the relevant process dimensions. However, you have shown that you can specify important process dimensions, and you would then be in the position to emphasize those strategies in your future career education efforts. Lastly, suppose that the unique contributions of both group membership and exposure variables are small relative to their co-funded or joint contributions. In this case, not only has the program succeeded in generating a treatment effect, but the education team has succeeded in specifying the particular process dimensions previously responsible for that effect. The Venn diagram presented in Figure 7 makes these relationships more clear.

## Questions of Relationships

This section covers those questions which concern the interrelationships among input, process, and output variables. These questions can vary considerably in their levels of complexity. A straightforward question about the relationship between any two continuous, normally distributed, variables can be answered by one of the most well known statistical analyses, the product-moment correlation. Such a question might be, "What is the relationship between students' family SES levels and their level of knowledge about careers?" On the other hand, a single question can address the tremendously complex interrelationship among several sets or domains of variables. For example, "What are the

concomitants to basic skills achievement among the independent *variable domains: teacher strategies, student exposure to career* education, and student background characteristics?" Such questions were formulated in Volume II, the Career Education Evaluation Final Report. To understand and to perform the data analysis techniques such as commonality analysis, canonical correlation, and factor analysis which are necessary to answer these questions, an extensive statistical background is required and sophisticated computer facilities must be available.

Most evaluators in the public schools in career education lack the appropriate statistical background and do not have the necessary computer "hardware" and "software" to do these more complex analyses. In such situations it would probably be best to formulate evaluation questions as parsimoniously as possible so that the more well known statistical analyses, such as product-moment correlation or chi square analysis, can be used.

## Questions of Prediction

Three subsets of questions will be discussed under this general category of questions of prediction. The first subset will concern common questions of prediction, such as, "What combination of career education strategies are most predictive of high gains in decisionmaking skills?" A second subset of questions will address the notion of causal dependency. For example, the question could be asked: "Does a positive self concept cause basic skills achievement or does achievement in the basic skills cause a positive self concept?" The third subset of questions addresses the effect of time: "Are gains in career education

outcome indicants constant over time, or do the benefits of career education diminish over time?" Another question might be: "Are outcome gains over time moderated by grade level?" Within this general category of inquiry, cohort design strategies will also be discussed. This is a method by which trends over time across grade levels can be examined.

Predicting to a Criterion

Many career educators desire some notion as to what combination of career education implementation strategies are best for achieving a particular outcome. Another question of considerable interest is: "Given my particular career education program, what input characteristics (e.g., sex, age, SES, etc.) are most predictive of high scores in a particular learner outcome?" The data analysis technique most commonly used to answer this form of question is regression analysis. Regression analysis is usually presented as a linear model for making predictions on a given criterion (or dependent variable) from one or more predictor (or independent) variables. The method yields a multiple correlation coefficient and a regression equation specifying a set of weights that can be used to optimize prediction of the criterion. In the event that the criterion variable is discreet rather than continuous (e.g., graduates who successfully obtained jobs versus those who were unsucessful), discriminant analysis would be the appropriate analysis. In both cases, the regression weights and discriminant weights give some clues as to the possible relative contribution of the various predictors to the observed variance in the criterion variable.

Causal Analysis

Some career education outcome evaluation questions relate to identifying causal dependencies among outcome variables. For example, is basic skill achievement dependent upon self-concept improvement or vice versa? When panel data is available with an appropriate lag, it is possible to address tentatively some of these questions, the answers to which might have immediate and profound impact on career education program development. A relatively new methodological tool, cross-lagged panel analysis, is available for analyzing longitudinal data with the intention of drawing causal inferences.

Let us assume that we have pre and post basic skills achievement and self-concept scores for 300 fifth grade students resulting in the matrix of correlations presented in Table 1 and graphically presented in Figure 8. For purposes of causal inference, the cross-lagged correlations ($r_{14}$,$r_{23}$) provide information of critical interest. Suppose that students' fall achievement caused the subsequent development of self concept (in spring), but that self-concept in the fall had no causal implications for later achievement. If such a causal pattern existed, we would expect to find the cross-lagged correlation between early achievement and later self-concept ($r_{14}$=0.70) to exceed significantly that relating early self-concept to later achievement ($r_{23}$=0.40). In fact, this finding is simulated in Table 1 and indicates the causal influence of achievement on self concept. Of course, we could have structured the table to indicate the opposite. Suppose that high self-concept heralded the development of achievement, but that early achievements held no necessary implications for later self-concept.

GRAPHIC PRESENTATION OF CROSS-LAGGED
PANEL DATA FROM TABLE 1

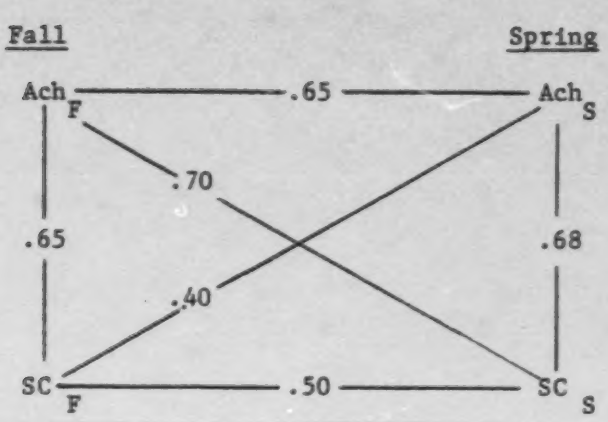



FIGURE 8

Table of Correlations Among Self-Concept
and Achievement Scores in Basic Skills
For a Sample of Fifth Graders

| EDS Science Test and Year | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Achievement - fall | - - - | | | |
| Self Concept - fall | .65 | - - - | | |
| Achievement - spring | .65 | .40 | - - - | |
| Self Concept - spring | .70 | .50 | .68 | - - - |

TABLE 1

In that case, $r_{23} > r_{14}$ would be the expected inequality. Of course, an absolute causal statement on the order of "X caused by Y" can not be supported on the basis of differences between cross-lagged panel correlations; however, inferences concerning the dominant causal relationship operating among self-concept and achievement variables would be legitimate and would enjoy the same logical status as those derived through the use of the experimental method. This method provides the analogue upon which this novel analytic approach is grounded.[3]

Trends Over Time

Graphic displays of data are often useful in uncovering or demonstrating predicted trends. Figure 9 is an illustrative example of career awareness scores across grade levels over a span of three years in an ongoing career education program beginning in 19761977. Also, for the purpose of this discussion, assume that this illustration represents a panel design in which the same students are surveyed over the span of three years.

You will note that the figure illustrates a program that has a constant and equal effect across grade levels. In 1977, the fifth grade scores exceeded the performance of the fourth grade by the same amount that the fourth grade exceeded the third grade. Also, all three groups of students improved a like

---

[3]Crano, William. "Causal Analysis of the Effects of Socio-economic Status and Initial Intellectual Endowment on Patterns of Cognitive Development and Academic Achievement." In The Aptitude-Achievement Distinction. Edited by Donald Ross. California Testing Bureau: McGraw-Hill, 1974.

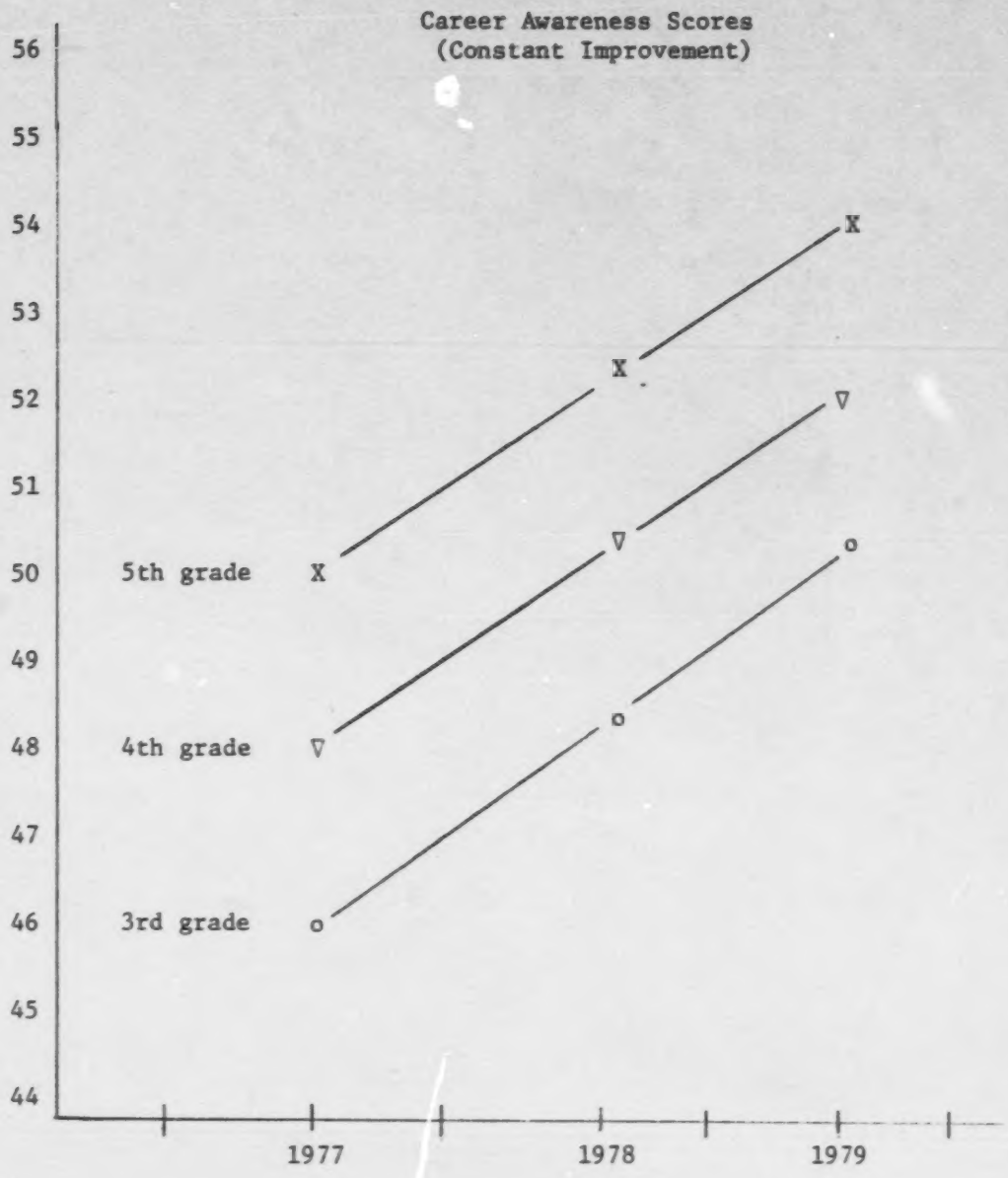

Career Awareness Scores
(Constant Improvement)
56
55
54
53
52
51
50
49
48
47
46
45
44
5th grade
4th grade
3rd grade
1977
1978
1979

FIGURE 9

amount over three years. This would be an example of where incremental improvement in career education is not taking place. If an incremental improvement were taking place, it would be expected that the performance of the fourth grade students in 1978 (the 1977 third graders) would exceed the level of performance achieved by the fourth grade students in 1977, and so forth.

Figure 10 depicts an entirely different situation. Improvement is neither constant nor equal. The results, comparing 1977 to 1978, demonstrate incremental improvements across all three grades. For example, the fourth graders of 1978 (the third graders of 1977) do outperform the fourth graders of 1977. Comparing the years 1978 and 1979, one observes what may be a "ceiling effect," that is, as the students move into junior high school their gains in career awareness begin to level off. It should also be noted that an "incremental improvement" was not demonstrated between 1978 and 1979. The fifth graders of 1979 (1977 third graders) actually performed at a lower level than the fifth graders of 1978. A similar but less dramatic loss in achievement level is shown for the sixth graders of 1979 (1977 fourth graders). It is hoped that Figure 10 demonstrates the importance of a longitudinal analysis. One can observe that the conclusions drawn from the illustrated program could be entirely different depending on the year and grade levels sampled.

Earlier reference was made to "cohort design strategies." This terminology usually refers to the use of one age group as a control group or baseline for younger students. Referring

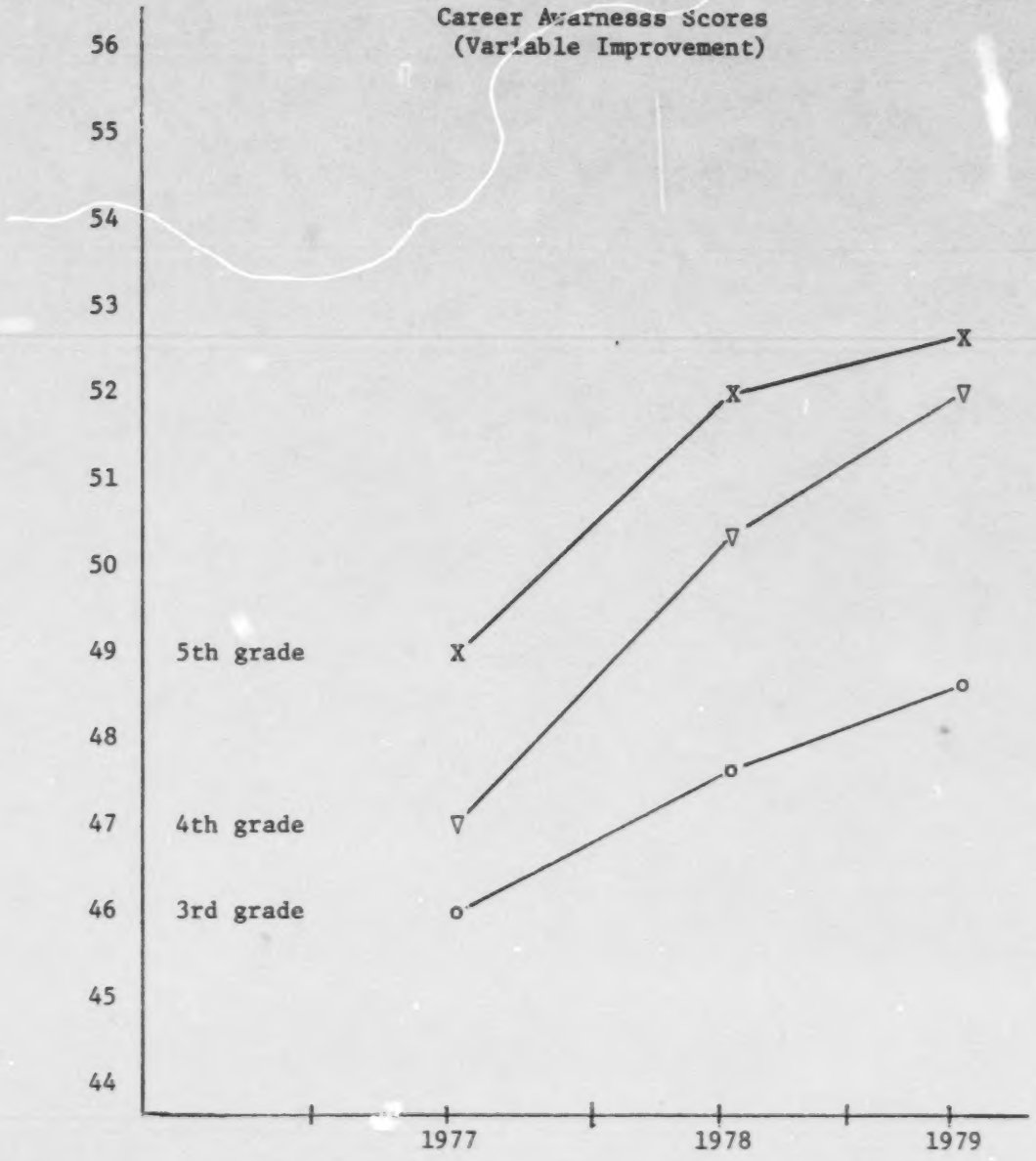

Career Awarnesss Scores
(Variable Improvement)
56
55
54
53
52
51
50
49
48
47
46
45
44
5th grade
4th grade
3rd grade
1977
1978
1979

FIGURE 10

to Figure 9, assume that the 1977 measures constitute a "befcre treatment" measure, and career education was implemented immediately after this pre-measure. On the basis of this illustration, one could conclude that the career education program had absolutely no effect on students' career awareness. For example, although the 1977 third graders did improve their scores, their level of performance did not exceed that of the fourth graders in 1977 who had not received career education. Figure 10, on the other hand, does demonstrate an effect of career education on 1977 third graders after having been exposed to the program for one year. Their performance as fourth graders exceeds that of their "cohort" by .5 standard deviations. It also should be noted that the above discussion will be applicable for those school systems utilizing a replicability design so long as the different groups of individuals at the various grade levels are shown to be similar regarding input variables.

# 9

## SPECIFY DATA COLLECTION AND REDUCTION PROCEDURES

Utilizing the list of instruments developed in Activity 6, decisions are made relative to how often and when each instrument will be administered. After these decisions are made, data collection and reduction procedures must be implemented. In order to implement the data collection and reduction procedures, a concise step-by-step procedure should be devised to facilitate each of the steps which are necessary to collect and reduce the data in the input, process, and outcome domains, regardless of when data is collected. Each of the steps is depicted in Figure 11 and is expanded in detail below.

I. Obtain Instruments

   A. Instruments commercially available.

      1. Obtain purchase order.

      2. Send purchase order.

      3. Store instruments on arrival.

   B. Instruments not commercially available.

      1. Consult federal guidelines, if appropriate.

      2. Consult professionals in the area of tests and measurement, if available.

      3. Design needed instruments.

FLOW CHART FOR THE COLLECTION OF
INPUT, PROCESS, AND OUTCOME DATA

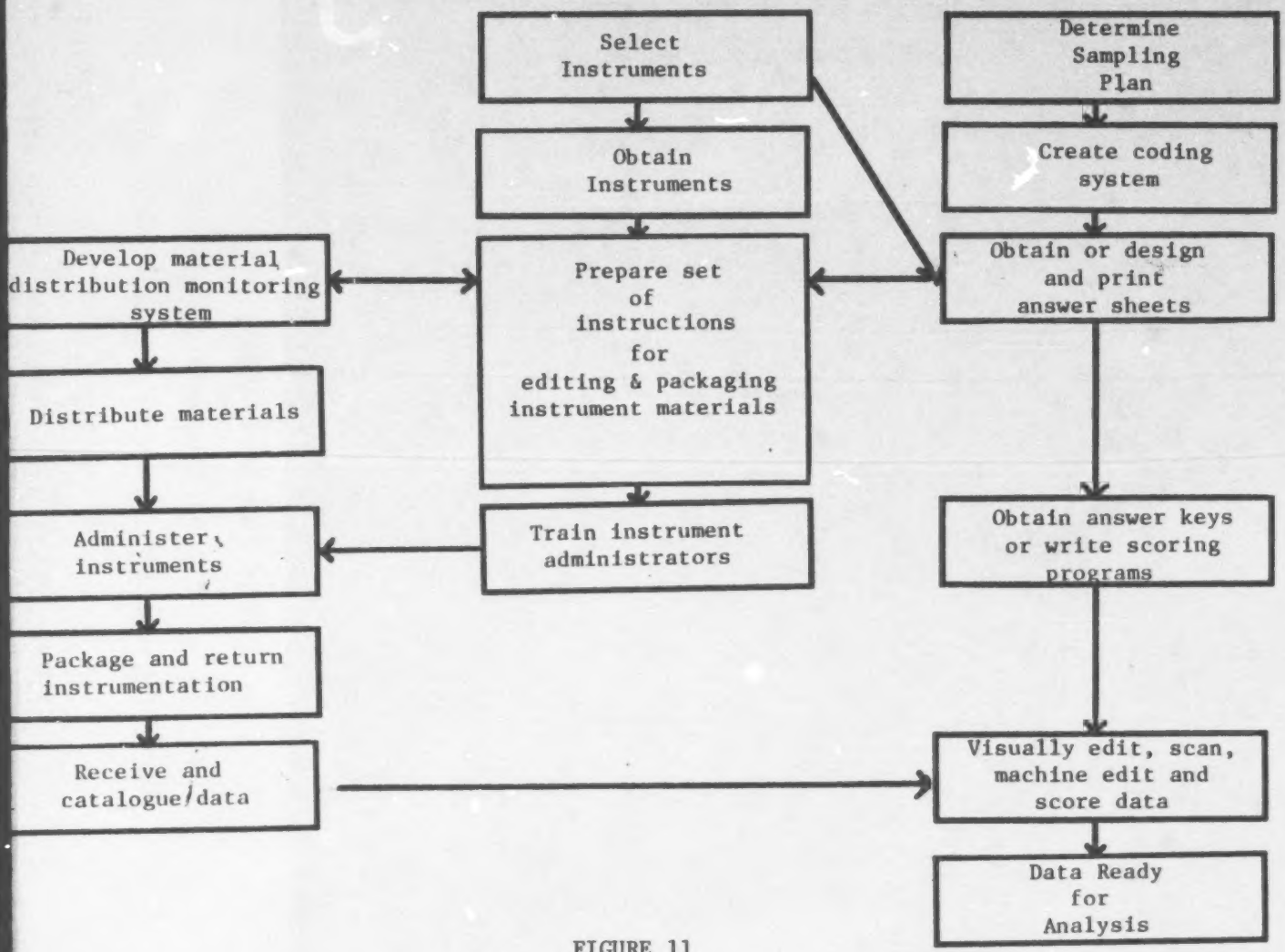

Select Instruments
Determine Sampling Plan
Obtain Instruments
Create coding system
Develop material distribution monitoring system
Prepare set of instructions for editing & packaging instrument materials
Obtain or design and print answer sheets
Distribute materials
Administer instruments
Train instrument administrators
Obtain answer keys or write scoring programs
Package and return instrumentation
Receive and catalogue data
Visually edit, scan, machine edit and score data
Data Ready for Analysis

FIGURE 11

4. Produce rough drafts of instruments.

5. Revise instruments.

6. Get proofs of revised instruments produced.

II. Prepare Set of Instructions For Editing and Packaging Instrument Materials.

A. Plan time line for data collection.

B. Determine instrument administrators.

C. Develop manual for instrument administrators.

1. Sample of each instrument.

2. Methodologies for administering, editing, and packaging instrument materials for return.

D. Produce manual for instrument administrators.

III. Train Instrument Administrators.

A. Schedule training of local instrument administrators.

B. Train local instrument administrators.

1. Background information.

a. Overview of the study including goals, objectivies, and background information on relevant legislation.

b. Objectives of the instrument administration.

2. Ethics of research.

a. Confidentiality of all information.

b. Anonymity of respondents.

c. The instrument administrators' responsibility for faithful reporting.

3. Field preparation.

   a. Positive attitudes.

   b. The setting for the instrument administration.

   c. Neat personal appearance.

   d. Making and keeping appointments.

   e. Voice appeal.

   f. Etiquette during instrument administration.

4. The approach.

   a. Introduction.

   b. Handling oneself with persons who are uncomfortable, suspicious, or otherwise reluctant regarding instrument administration.

   c. Gaining the respondent's cooperation.

5. Techniques of administration

   a. Instrument security.

   b. Instrument distribution.

   c. The problem of defective instruments.

   d. General administrative duties.

6. Simulations of instrument administration.

IV. Create Coding System.

A. School district.

1. Assign ID number to each school tested in the district.

2. Assign ID number to each test administrator in each school.

3. Assign ID number to each teacher whose students are tested in each school.

4. Assign ID number to each grade (01-12), class period (01-10).

5. Assign ID number to treatment and companion groups, if appropriate.

V. Obtain or Design and Print Answer Sheets.

A. Purchase commercially available answer sheets - or design individual answer sheets.

B. Design headers compatible with answer sheets.

C. Get answer sheets produced.

D. Collate and staple answer sheets into response booklets.

E. Assemble response booklets in envelope packages.

F. Give instructions on how answer booklets and envelopes are to be returned.

VI.. Obtain Answer Keys or Write Scoring Programs.

A. Develop scoring programs for each instrument or obtain answer keys.

B. Develop dummy data set(s).

C. Write programs.

D. Test programs on dummy data set(s).

E. Revise programs, where necessary.

VII. Develop Materials Distribution Monitoring System.

A. List all materials being distributed.

B. Develop list of location addresses and contact persons with phone numbers.

C. List quantity of each of materials to be sent to each location.

D. Record expected return date of materials.

VIII. Distribute Materials.

A. Determine most efficient method of distribution. (Example: Organize materials by administrator.)

B. Package materials according to distribution monitoring system.

C. Distribute materials.

IX. Administer Instruments.

A. Schedule each set of respondents for each instrument.

1. Location each day.

2. Time each day.

B. Check materials.

1. Correct set of instruments.

2. Correct numbers of each instrument and response booklets.

3. Appropriate administration manuals.

C. Congregate respondents in appropriate location at designated time.

D. Administer instruments using procedures developed in training session.

E. Collect instruments.

F. Check for correct number of instruments and response booklets, as well as for proper identification on each.

G. Record and report any anomalies in test situation.

X. Package and Return Instrumentation.

A. Check for correct numbers of instruments and response booklets by subgroup.

B. Group instruments by primary groupings specified in administration manual.

C. Group response booklets by primary groupings specified in administration manual.

D. Store instruments appropriate for future use.

XI. Receive and Catalogue Data.

A. Receive materials.

B. Check for appropriateness of quantity and location.

C. Record quantity and location.

XII. Visually Edit, Scan, Machine Edit, and Score Data.

A. Manually processed data.

1. Visually edit to assure proper ID information.

2. Handscore forms.

3. Verify scoring, if feasible, by second scoring.

4. Compile score rosters.

5. Store answer sheets.

B. Machine processed data.

1. Visually edit to assure proper ID information.

2. Scan optically scannable forms or keypunch data.

3. Store processed forms.

4. Store data on computer.

5. Machine score data.

6. Print score rosters.

# 10

## SPECIFY REPORTING REQUIREMENTS AND PROCEDURES

The format, content, and style of evaluation reports are determined by the material to be included, the audience(s) to whom the reports are directed, and the resources, particularly time, available for their preparation. This is true whether the report is of the interim "progress" variety or is a detailed explication of evaluation results.

### Special Concerns

Three specific concerns need to be kept in the forefront of those writing the reports: the reports should be readable; they should focus on the information needs of the primary target audience; and they should provide information at programmatically useful times. The target audience will usually be those entrusted with policy level decisionmaking. These three concerns overlap, of course, but all three are important.

For reports to be readable, unduly complex and/or technical material should be relegated to appendices or footnotes. You might impress some with technical wizardry, but they may not understand the results. Your audience should determine your approach here; if they expect sophistication, then provide it. If not, the efforts to provide it may be counter-productive.

Nearly all evaluations provide ample areas for discussions of topics that are fascinating to the evaluators. Unless these are also fascinating to the reader, they may remain unread; further, in avoiding them, the reader may miss that which is important. Useful techniques for accomodating both are the use of an executive summary highlighting the results, brief summaries for each chapter hitting the most salient points, and a "users guide" in the introduction explaining to readers the structure of your report.

The reports should be timely. It may do program staff little good to receive an evaluation report one month after they have to reapply for funding. Of course, you may not be able to mesh your schedules as completely as you desire, but you may be able to meet such program needs informally.

S ecifying Report Requirements

The following are a brief set of suggestions for accomplishing this activity.

- Analyze audience size and characteristics. You may find you have different audiences. If so, the need to include different levels of information in different sections of your report will be especially important.
- Specify reporting format. This requires the develcpment of outlines for each of the reports requested of you. Following development, the draft outlines should be submitted to program staff to ensure coverage of the topics of interest to them.

- Identify the frequency and due dates of reports. In the initial stages of an evaluation, all those involved desire an informational bonanza. By the conclusion of the effort, this bonanza has often become played out. Interim reports can prove useful, but not if they are so frequent as to take excessive time or if they are too infrequent to allow discovery of potential problems.

In addition, it is expecially important to allow sufficient final report writing time following receipt, reduction, and analysis of the data.

# 11

## IDENTIFY MAJOR ASSUMPTIONS AND RESTRAINTS

Very seldom are there evaluations which are "pure" and devoid of any constraints or assumptions. Though the types of assumptions and constraints are numerous, it is useful to divide them into two general varieties: those which operate internally to the project and those which exist outside the project but still affect it.

The first of these types is more significant, not because its import is necessarily greater, but because these internal constraints and assumptions are more likely to be controllable by the evaluation team. Such internal constraints as the time available for different activities, the money available to purchase equipment, supplies, and personnel, and the materials and facilities to be used can all have major effects on your evaluation. No general guidelines can be offered because of the various forms these constraints can take except to reiterate the need to be aware that they potentially all exist.

External constraints are not, by their very nature, open to much manipulation by you or your staff. Several types of these would include the political, social, and economic context in which

you are working and the procedural and methodological skills open to you.

## Completing This Activity

In completing this activity you should attempt to identify any complications which have acted as limitations on the design. To do this, NTS suggests:

- Review each activity in order to
- Identify the internal constraints and assumptions and to
- Identify the external constraints and assumptions so that you can
- Estimate the impact of these factors.
- Produce a concise statement of the estimated impact and share it openly with the program staff before the conclusion of the design process.

# 12

## DEFINE SUPPORT REQUIREMENTS

Support requirements are necessary for all evaluations. The amount and types of support are dependent upon the scope and complexity of the evaluation. All materials and any service supplied by an individual or agency other than the evaluation team are considered support. The types of support requirements ne ed can be determined by reviewing each activity in the flow chart and asking, "What is needed to complete this activity?" Support requirements generally fall into four categories:

1. People (e.g., data collectors, test scorers, etc.)
2. Services (e.g., key punching, computer analysis, etc.)
3. Materials (e.g., instruments, supplies, etc.)
4. Facilities (e.g., housing, storage, etc.)

To Complete This Activity:

- Define support requirements for each activity.
- Identify available support.
- Identify associated costs.

- Reconcile required support with available support.
- Develop procedures for securing needed support.

The product of these process steps is a plan for securing the needed support to complete the evaluation.

# 13

## DEVELOP SCHEDULE OF EVALUATION ACTIVITIES

Having developed the evaluation questions and designed the other activities which are to take place in your evaluation, it becomes necessary to schedule these activities to allow their completion in a manner which will accomplish the goals of the evaluation. Several scheduling techniques are available. These include using the critical path method, a Gantt chart, or PERT. These are sophisticated techniques which have been developed for the completion of complex or large scale activities in both industry and government. However, for a smaller school sytem they involve time and effort which might be better spent in the evaluation.

The scheduling of evaluation is an effective tool for planning as well as a quality control mechanism once evaluation is operational. To maximize the utility of the evaluation schedule, it should be viewed as a flexible document amenable to changes which will allow for unforeseen situations, yet rigid enough to specify exact dates for the completion of activities as listed and to insure that no major milestone has been omitted.

A schedule of activities is potentially useful because it delineates the major evaluation milestones and allows for their completion in a coherent manner. Among the more common evaluation milestones are the following:

1. Instrument construction and pilot testing.
2. Conferences with building instructors.
3. Instrument administration.
4. Data reduction.
5. Data analysis.
6. Report preparation and submission.

In general, the schedule should include what, when, how, who, where, and why: what is the activity; when should it be completed; how should it be completed; who should complete it; where it should be completed; and why it should be completed.

For the completion of Activity 13, these are the major elements that have to be considered:

1. The identification of major tasks or activities for the evaluation.
2. The identification of completion dates for each task using report dates as reference points.
3. The building of a scheduling activity using any scheduling techniques with which the evaluation team feels comfortable.

The product of this activity is a schedule of evaluation activities sequenced in time with completion dates and responsible individuals identified.

# 14

## LIST BUDGET REQUIREMENTS

Any evaluation undertaken is limited by several factors; these include time, money, materials, and local restrictions and regulations. These limiting factors act as constraints on the quality and scope of an evaluation. In this regard, the development of a budget for evaluation activities proves an important activity in the design process. Budgeting often results in changes in the evaluation design as questions and processes are eliminated because of the cost factor.

In practice, you, as project manager, will generally have a figure in mind as available for conducting the evaluation. This approximate figure is generally agreed upon before the evaluation design is formulated. the most cost efficient use of these funds must be determined by you, the members of your evaluation team, and the major users of the evaluation results. The deletion or addition of any activities should be the result of a shared decisionmaking process involving both the members of the design conference and your evaluation team.

Thus, use of evaluation dollars is dependent upon information priorities established in Activities 2 and 5. It is a rare occurrence when the evaluation dollar available to you is sufficient to meet all the identified information needs in Activity 2. Therefore, care should be taken in assigning the limited evaluation resources so that maximal benefits, in terms of useful information for decision makers, can be realized. Several budget categories which should be considered in estimating evaluation expenditures include the following aspects of overhead costs:

Fringe

- Mandatory (FICA, unemployment, workmen's compensation)
- Annual Leave
- Sick Leave
- Holidays
- Health and Life Insurance

General Overhead

- Rent
- Furnishings
- Utilities
- Office Supplies
- Xerox
- Equipment (Hazeltine, dictating, typewriters)
- Telephone
- Postage

- Depreciation
- Recruitment/Relocation (general)
- Data Processing Maintenance
- Staff Development

General and Administration

- Bookkeeper
- Administrative Salaries
- Financial and Administration

The budget requirement phase in Activity 14 can be facilitated by seeking out the knowledgeable personnel in your school system. A procedure might already be available to you which will make your job much easier. It will also insiure that any budget that is developed as a result of this activity will be in conformance with the guidelines set down by your district. Consultation with the controller's office will also insure that realistic figures are developed based upon past history for any given category.

In summation, several suggested process steps for completing Activity 14 are suggested below:

1. Estimate the manpower requirements including professional, junior, clerical, and casual labor.
2. Estimate salary and wages for each class of staff identified in the previous line.
3. Review each evaluation activity and enter estimates in each category provided above.
4. Compare budget estimates with available resources.
5. Review budget of evaluation activities required to meet budget constraints.
6. Review budget with district financial personnel.

The product of these foregoing activities will be a line item budget of proposed expenditures for evaluation activities to be undertaken.

# 15

## DEVELOP MANAGEMENT PLAN

A management plan should be formulated to ensure the proper development and implementation of the evaluation plan and to utilize properly person power, materials, and resources identified as necessary for completion of this activity. The structure of the management plan should be coincident with the scope and complexity of the evaluation. This activity is the culmination of Activity 14 and should flow as a natural result from the activities which have preceded it.

A useful methodology for bringing all the previous activities together in a coherent management plan might be the implementation of management by objectives (MBO). In this management methodology the project manager and the evaluation team meet during the course of the evaluation and confer to insure that the activities and processes detailed in Activity 13 are being performed according to schedule. MBO is particularly advantageous for this phase of the evaluation, as two of the grounding factors for MBO are constant feedback and subsequent reassessment of objectives and goals. The use of MBO is only a suggestion; you may wish to use another management plan more in line with the modus operandi of

your system. Regardless, the management plan should deal with staff assignments, internal reporting, conference scheduling, general policies and procedures, budget control procedures, and the like.

The project staff developing the management plan should be cognizant of two vital factors. The evaluation effort should interface successfully with the program management of the career education program, and the paths of communication between various staff and line personnel should be identified with roles and relationships explained. The evaluation should be planned to avoid any disruption of the general goals of the career education program that it is evaluating. The evaluation should not, for example, plan an activity in a school on a day in which students at that school are scheduled for a field trip.

The importance of developing a project management plan for the evaluation cannot be overly stressed, because an evaluation can fail or succeed depending upon the effectiveness of its management plan. School personnel can at times view evaluation as an invading evil rather than as a tool for helping them perform their jobs in a more efficient manner.

In summation, the implementation of a coherent management plan consists essentially of two steps:

1. The identification and assignment of staff to various evaluation activities.

2. The development of an organizational chart which can be understood by all members of the evaluation team, program staff, and school staff so that both the program and the evaluation can be efficiently run.

The product of these foregoing steps is a document which delineates the roles and relationships within the evaluation, the program, and the school system. The product of this activity is an organizational chart which should be widely published or posted.

# 16

## PREPARE THE EVALUATION DESIGN DOCUMENT

The evaluation design document is the plan, structure, and strategy for the evaluation of a program. It provides the overall scheme for the entire evaluation effort and includes a complete discussion of tasks and considerations composing the evaluation activity. The design document is analogous to a blueprint; it specifies the product, and the components and activities comprising that product. Without structure--without a sound purpose, design, and plan--the evaluation is empty, and little of value can be expected, much less realized.

Both the experience gained in developing the evaluation design and the evaluation document itself are of considerable utility. Developing the design document involves interaction between program managers and program evaluation staff; thus, the development process facilitates much needed communication. A very useful tool for formulating the evaluation is the initiation of an evaluation planning conference involving the program evaluation team, the program management staff, and representatives from program operations staff (teachers, paraprofessionals, etc.).

Such a conference can be used to enlighten all involved in the complexities of evaluation, the need for staff support, and the information to be derived from the evaluation activities.

The evaluation design document serves a variety of functions including the following:

1. It provides a vehicle for ascertaining the internal integrity and logical consistency of the evaluation effort.
2. It provides a systematic set of procedures and specifications for the evaluation to follow.
3. It provides documentation for agreed-upon activities, roles, and relationships, thereby minimizing misunderstandings.
4. It provides a control document whereby project management can monitor the evaluation activities.
5. It serves as an excellent internal communication vehicle and external dissemination device.
6. It fulfills state and federal guideline requirements for proposals and/or program designs.

The evaluation design document includes a series of sections or chapters, each of which describes the product of certain activities in the design flow chart. The following pages describe the various sections of a comprehensive evaluation design document.

## Preliminary Materials

The design document includes several pages of important information prior to the first chapter. The title page, which appears first, contains as a minimum the following information: (1) Title of the Program; (2) Authors of the Document; (3) Recipient of the Document; and (4) The Submission Date. The next page is usually a letter of transmittal acknowledging

## Table of Contents For An Evaluation Design Document

Chapter Contents

The following discussion deals with the contents of each chapter in the evaluation design document.

Chapter 1. Program Summary - This chapter should provide an overview of the program activities and procedures. The statement should be concise and provide the reader with an understanding of the scope and intent of the program.

Chapter 2. Program Objectives - A list of the program objectives is sufficient for completing this chapter.

Chapter 3. Conceptual Framework - This chapter should include a statement of the evaluation team's philosophy or conception of the role of evaluation. If a particular evaluation model is being used, then a brief description of the model should be included.

Chapter 4. Evaluation Questions - The specific evaluation questions should be listed. If a large number of questions are being investigated, then they should be logically grouped according to audience or orientation.

Chapter 5. Data Base Elements - Each data element should be identified. If the data elements are organized into files, then the file structure should be described.

Chapter 6. Instrumentation - Each instrument to be used on the project should be listed along with a brief description of the intent of the instrument.

Chapter 7. Population Selection Procedures - The technique for selecting the samples for evaluation purposes should be described. Comparability between treatment and control groups (if applicable) should be discussed.

Chapter 8. Data Analysis Techniques - The specific data analysis techniques to be used in answering each evaluation question should be listed and discussed.

Chapter 9. Data Collection and Reduction Procedures - The plans for collecting data and administering instruments should be described. A discussion of the method to be used for data reduction should be provided.

Chapter 10. Reporting Requirements and Procedures - All reports and report formats should be identified and the submission dates listed. If some reports have different audiences, these should be discussed.

Chapter 11. Major Assumptions and Constraints - This chapter should identify and discuss major assumptions and constraints with empahsis on how these considerations will impact the evaluation effort.

Chapter 12. Support Requirements - List and describe the needed support requirements. Such considerations as access and cost should be discussed.

Chapter 13. Activities Schedule - This chapter should include some type of scheduling plan. As a minimum, the schedule should include the activity to be completed, the completion date, and the individual responsible for completing the activity.

Chapter 14. Management Plan - This chapter should include an organizational chart and a description of evaluation staff.

Chapter 15. Estimated Budget - This chapter consists of a line item budget statement for all evaluation related costs on the project.

Usually, it is a good idea to include in an appendix instruments which have been locally developed. Another useful appendix is the summary of any evaluation design conference which might shed some light on the developmental process through which the evaluation design has progressed.

FIGURE 12

the obligation, contractual or otherwise, under which the evaluation design was completed. The next page(s) is the Table of Contents which should include major chapter headings. The Table of Contents is usually followed by a listing of tables and figures.

Figure 12 presents an example of a table of contents for an evaluation design document.

# APPENDIX

## EVALUATION DESIGN CHECKLIST

FIGURE 13

EVALUATION DESIGN CHECKLIST

| ACTIVITY Steps | PERSON RESPONSIBLE | DATE TO BE COMPLETED | DATE ACTUALLY COMPLETED | COMMENTS |
|---|---|---|---|---|
| 1. REVIEW PROGRAM DESIGN, OBJECTIVES AND RATIONALE | | | | |
| a. Review rationale | | | | |
| b. Establish congruence between program objectives and program rationale | | | | |
| c. Establish that program objectives are operationally stated | | | | |
| d. Establish congruence between program objectives and program activities | | | | |
| 2. SPECIFY INFORMATION REQUIREMENTS AND MANAGEMENT NEEDS | | | | |
| a. List all potential information users | | | | |
| b. List all information domains | | | | |
| c. Create matrix contrasting users against domains | | | | |
| d. Check which users need information in each domain | | | | |
| e. Determine most essential users and domains | | | | |
| f. Create "program information requirements matrix" | | | | |
| 3. IDENTIFY CONCEPTUAL FRAMEWORK FOR EVALUATION | | | | |
| a. Review available evaluation models | | | | |
| b. Review information requirements | | | | |
| c. Synthesize a summary of the purpose and intent of the evaluation | | | | |

Evaluation Design Checklist
(continued)

| ACTIVITY Steps | PERSON RESPONSIBLE | DATE TO BE COMPLETED | DATE ACTUALLY COMPLETED | COMMENTS |
|---|---|---|---|---|
| d. Summarize program information mandate | | | | |
| 4. SPECIFY EVALUATION QUESTIONS OF INTEREST | | | | |
| a. Review program information matrix | | | | |
| b. Conceptualize evaluation questions | | | | |
| c. Negotiate final evaluation questions | | | | |
| d. Develop procedures for adding new evaluation questions | | | | |
| 5. SPECIFY DATA BASE STRUCTURE AND CONTENTS | | | | |
| a. "Map" phrases in each evaluation question | | | | |
| b. List elements which will define each phrase | | | | |
| c. Eliminate redundant data elements | | | | |
| d. Create data files by listing data elements by audiences and actions | | | | |
| e. Select data base structure | | | | |
| 6. SELECT INSTRUMENTATION | | | | |
| a. List most appropriate sources of information | | | | |
| b. Review available data collection instruments | | | | |
| c. Select from existing instruments or develop your own | | | | |

Evaluation Design Checklist
(continued)

| ACTIVITY Steps | PERSON RESPONSIBLE | DATE TO BE COMPLETED | DATE ACTUALLY COMPLETED | COMMENTS |
|---|---|---|---|---|
| 7. SPECIFY POPULATION SELECTION PROCEDURES | | | | |
| a. List sample sizes needed to respond to each evaluation question | | | | |
| b. Choose most suitable sampling plan | | | | |
| 8. SPECIFY DATA ANALYSIS TECHNIQUES | | | | |
| a. List alternative techniques for each evaluation question based on typed data and sample | | | | |
| b. Establish the appropriateness of the statistical assumptions for each alternative | | | | |
| c. Determine ease and use of interpretation of each alternative | | | | |
| d. Select data analysis technique for each question | | | | |
| 9. SPECIFY DATA COLLECTION AND REDUCTION PROCEDURES | | | | |
| a. Establish frequencies of instrument administration | | | | |
| b. Select most appropriate reduction procedures based upon data base structure and the analyses planned | | | | |
| 10. SPECIFY REPORTING REQUIREMENTS AND PROCEDURES | | | | |
| a. Analyze audience size and characteristics | | | | |
| b. Specify reporting format | | | | |
| c. Identify frequency and due dates for reports | | | | |

Evaluation Design Checklist
(continued)

| ACTIVITY Steps | PERSON RESPONSIBLE | DATE TO BE COMPLETED | DATE ACTUALLY COMPLETED | COMMENTS |
|---|---|---|---|---|
| 11. IDENTIFY MAJOR ASSUMPTIONS AND CONSTRAINTS | | | | |
| a. Review each activity | | | | |
| b. Identify internal constraints and assumptions | | | | |
| c. Identify external constraints and assumptions | | | | |
| d. Estimate impact of these constraints and assumptions | | | | |
| 12. DEFINE SUPPORT REQUIREMENTS | | | | |
| a. Define support requirements for each activity | | | | |
| b. Identify available support | | | | |
| c. Identify associated costs | | | | |
| d. Reconcile required with available support | | | | |
| e. Develop procedures for securing needed support. | | | | |
| 13. DEVELOP SCHEDULE OF EVALUATION ACTIVITIES | | | | |
| a. Identify tasks and activities | | | | |
| b. Identify completion dates for activities using report dates as reference points | | | | |
| c. Develop a schedule of activities using any scheduling technique with which the evaluation team feels comfortable | | | | |

Evaluation Design Checklist
(continued)

| ACTIVITY Steps | PERSON RESPONSIBLE | DATE TO BE COMPLETED | DATE ACTUALLY COMPLETED | COMMENTS |
|---|---|---|---|---|
| 14. LIST BUDGET REQUIREMENTS | | | | |
| a. Estimate person-power requirements by class of staff | | | | |
| b. Estimate salaries and wages by class of staff | | | | |
| c. Review each activity and estimate costs | | | | |
| d. Compare estimates with available resources | | | | |
| e. Revise budget of evaluation activities as required to meet budget constraints | | | | |
| 15. DEVELOP MANAGEMENT PLAN | | | | |
| a. Identify and assign staff to evaluation activities | | | | |
| b. Define relationships between evaluation team and program staff | | | | |
| c. Develop organizational chart | | | | |